

-1251-

PI +LGLS+F+ +L+ L +K +V++ K + F L+W A+ E RKQSILKF
 Sbjct: 121 PIFRRRLGLSLFIFIIIVLILLALKRVVLSRKTRYFLRGNRDWAQAVAFESNRKQSILKF 180

Query: 181 FSLFTNVKGISTSVKRRSFLDGILKLISKTPSRLWTNLFVRAFLRSSDYLGLTIRLVTLN 240
 +SLFT VKGIST VK R++L+ +LKL+ +TPS LW +L+ RAFLRSSDYLG +RL+ L+
 Sbjct: 181 YSLFTTVKGISTKVKERTYLNPLLKLVKQTPSNLWLSLYARAFLRSSDYLGFLRLMLLS 240

Query: 241 ILSVIFVNETYIALALAFVFNLYLLLFQLLALGHFFDYQYMNQLYPVRLNAKASQLKGFRLR 300
 LSV F++ YL+++LA +FNLY++FQLL+L +H+DY YM LYP +K + FLR
 Sbjct: 241 SLSVFFIHNLYLSVSLALIFNLYLVVFQLLSLYYHYDYHYMTSLYPENSRSKKKNMLSFLR 300

Query: 301 VLSYAVTVIDSILIRELKPVILLIVMLIVTEYYIPYKIKKMID 344
 LS+ + +++ + ++LIV M+ + Y+PYK+KK+ID
 Sbjct: 301 GLSFLMLIVNMLCCSSAPKALILIVGMVFIACIYLYPYKLKKIID 344

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1120

A DNA sequence (GBSx1195) was identified in *S.agalactiae* <SEQ ID 3463> which encodes the amino acid sequence <SEQ ID 3464>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2821(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00284 GB:AF008220 YtmP [Bacillus subtilis]
 Identities = 69/214 (32%), Positives = 121/214 (56%), Gaps = 1/214 (0%)

Query: 12 PLRGKSGKAYIGTYPNGERVFKYNTTPILPALAKEQIAPQLLWARRTSNGDMMSAQEWL 71
 P G +G AY + NG+++F+K N++P L L+ E I P+L+W +R NGD+++AQ W+
 Sbjct: 20 PAGGATGDAYYAKH-NGQQFLKRNSSPFLAVLSAEGIVPKLVWTKRMENGDVITAQHW 78

Query: 72 DGRTLTLEDMSGKQIIHILLRLHKSRLPLVNQLLQGLGYKIENPYDLLMDWEKQTPIQIREN 131
 GR L +DM + + +L ++H S+ L++ L +LG + NP LL ++ + +
 Sbjct: 79 TGRELKPKDMSGRPVAELLRKIHTSKALLDMLKRLKEPLNPGALLSQLKQAVFAVQSS 138

Query: 132 TYLQSIIVTELKRSLEPFRTVEVATVHGDIKHSNWWITTSGLIYLVWDWSVRLTDRMYDVA 191
 +Q + L+ L E + H D+ H+NW+++ +YL+DWD + D D+
 Sbjct: 139 PLIQEGIKYLEEHLHEVHFGEKVVCHCDVNHNNWLLSEDNQLYLIDWDGAMIADPAMD LG 198

Query: 192 YILSHYIPQKHWDLSYGYKDNEKVVWSKIIWY 225
 +L HY+ + W+ WLS YG + E + ++ WY
 Sbjct: 199 PLYHYVEKPAWESWLSMYGIELTESLRLMAWY 232

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3465> which encodes the amino acid sequence <SEQ ID 3466>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2686(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1252-

Identities = 214/262 (81%), Positives = 242/262 (91%)

Query: 1 MTISNQELTLTPLRGKSGKAYIGTYPNGERVFKYNTTPILPALAKEQIAPQLLWARRTS 60
 +T + QELTLTPLRGKSGKAY GTYPNGE VF+K NTPILPALAKEQIAPQLLWA+R
 Sbjet: 1 VTTTEQELTLTPLRGKSGKAYKGTYPNGECVFIKLNTPILPALAKEQIAPQLLWAKRMG 60

Query: 61 NGDMMSAQEWLDGRTLTKEDMGSKQIIHILLRLHKSRLVNQLQLGYKIENPYDLLMDW 120
 NGDMMSAQEWL+GRTLTKEDM SKQIIHILLRLHKS+ LVNQLQL YKIENPYDLL+D+
 Sbjet: 61 NGDMMSAQEWLNGRTLTKEDMNSKQIIHILLRLHKS KKLNVNQLQLNYKIENPYDLLVDF 120

Query: 121 EKQTPIQIRENTYLSIVTELKRSLEPFRTVATVHGDIKHSNWVITTSGLIYLVWDWS 180
 E+ P+QI++N+YLO+IV ELKRSLEPF++EVATVHGDIKHSNWVITTS+I+LVWDWS
 Sbjet: 121 EQNAPLQIQNSYLAIVKELKRSLEPFKSEVATVHGDIKHSNWVITTS+I+LVWDWS 180

Query: 181 VRLTDRMYDVAYILSHYIPQKHWDWLSYGYKDNEKVSIIWYQFSYLSQIIKCFDK 240
 VRLTDRMYDVAY+LSHYIP+ W +WLSYGYK+N+KV KIIWYQFS+L+QI+KCFDK
 Sbjet: 181 VRLTDRMYDVAYLLSHYIPSRWSEWLSYGYKNDKVMQKIIWYQFSHLTQILKCFDK 240

Query: 241 RDMEHVNQEYELRKFRELIK 262
 RDMEHVNQEY LRFRE+ +K
 Sbjet: 241 RDMEHVNQEYALRKFREIFRK 262

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1121

A DNA sequence (GBSx1196) was identified in *S.agalactiae* <SEQ ID 3467> which encodes the amino acid sequence <SEQ ID 3468>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4529(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00285 GB:AF008220 YtmQ [Bacillus subtilis]
 Identities = 126/211 (59%), Positives = 161/211 (75%)

Query: 1 MRVRKRKGAEHLENNPHYVISNPPEAKGRWHEIFGNNNPIHIEVSGSGKAFITGMAEQN 60
 MR+R + A++ L N ISNP + KG+W+ +FGN+NPIHIEVG+GKG FI+GMA+QN
 Sbjet: 1 MRMRHKPWADDFLAENADIAISNPADYKKGWNTVFGNDNPIHIEVGTGKGQFISGMAKQN 60

Query: 61 PDINYIGIDIQLSVLSYALDKVLDGSAKNIKLLLDGSSLSNYFDTGEVDLMYLNFSDPW 120
 PDINYIGI++ SV+ A+ KV DS A+N+KLL +D +L++ F+ GEV +YLNFSDPW
 Sbjet: 61 PDINYIGIELFKSVIVTAVQKVKDSEAQNKLNLNIDADTLTDVFEPGEVGRVYLNFSDPW 120

Query: 121 PKKKHEKRRITYKTFLDTYKDLPEQGEIHFKTDRGLFEYSLASFSQYGMTLKQVWL 180
 PKK+HEKRRITY FL Y++++ + G IHFKTDNRGLFEYSL SFS+YG+ L V LDL
 Sbjet: 121 PKKRHEKRRITYSHFLKYYEVMGKGSIHFKTDNRGLFEYSLKSFSEYGLLLTYVSLDL 180

Query: 181 HASDYQONIMTEYERKFSNKGQVIYRVEARF 211
 H S+ + NIMTEYE KFS GQ IYR E +
 Sbjet: 181 HNSNLEGNIMTEYEEKFSALGQPIYRAEVEW 211

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3469> which encodes the amino acid sequence <SEQ ID 3470>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

-1253-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3303(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 179/211 (84%), Positives = 193/211 (90%)

10

Query: 1 MRVRKRKGAEHLENNPHYVISNPEEAKGRWHEIFGNNNPIHIEVSGSGKAFITGMAEQN 60
 MRVRKRKGAEHL NNPHYVI NPE+AKGRWH++FGN+ PIHIEVSGSGK FITGMA +N
 Sbjct: 1 MRVRKRKGAEHLANNPHYVILNPEDAKGRWHDVFGNDRPIHIEVSGSGKGFITGMALKN 60

15

Query: 61 PDINYIGIDIQLSVLSYALDKVLDSGAKNIKLLLDGSSLSNYFDTGEVDLMYLNFSDPW 120
 PDINYIGIDIQLSVLSYALDKVL S N+KLL VDGSSL+NYF+ GEVD+MYLNFSDPW
 Sbjct: 61 PDINYIGIDIQLSVLSYALDKVLASEVPNVKLLRVDGSSLTNYFEDGEVDMMYLNFSDPW 120

20

Query: 121 PKKKHEKRRLTYKTFDITYKIDILPEQGEIHFKTDNRGLFEYSLASFSQYGMTLQVWLDL 180
 PK KHEKRRLTYK FLDITYK ILPE GEIHFKTDNRGLFEYSLASFSQYGMTL+Q+WLDL
 Sbjct: 121 PKTKHEKRRLTYKDFLDITYKRILPEHGEIHFKTDNRGLFEYSLASFSQYGMTLRQIWLDL 180

Query: 181 HASDYQQNIMTEYERKFSNKGQVIYRVEARF 211
 HAS+Y+ N+MTEYE KFSNKGQVIYRVEA F
 Sbjct: 181 HASNYEGNVMTYEYEEKFSNKGQVIYRVEANF 211

25

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1122

A DNA sequence (GBSx1197) was identified in *S.agalactiae* <SEQ ID 3471> which encodes the amino acid sequence <SEQ ID 3472>. Analysis of this protein sequence reveals the following:

30

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35

bacterial cytoplasm --- Certainty=0.1311(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40

>GP:BAB06136 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 61/124 (49%), Positives = 81/124 (65%), Gaps = 2/124 (1%)

45

Query: 2 GGDYVLSILIDKPGGITVEDTAQLTDVVSPLLDITQDPDFPEQYMLEVSSPGLERPLKTA 61
 G D+ L + ID G+ +ED +++++ +S LD + DP + Y LEVSSPG ERPLK
 Sbjct: 33 GKDWFLRVFIDSETGVLDLEDCKGVSERLSEKLD--ETDPIEQAYFLEVSSPGAERPLKRE 90

Query: 62 EALSNAVGSYINVSLYKSIDKVKIFEGDLLSFDGETLTIDYMDKTRHKTVDIPYQTVAKA 121
 + L ++G ++V+LY+ ID K EG+L FDGETLTI+ KTR KTV IPY VA A
 Sbjct: 91 KDLRSIGKNVHVTLYEPIDGEKALEGELTEFDGETLTIEIKIKTRKKTVTIPYAKVASA 150

50

Query: 122 RLAV 125
 RLAV
 Sbjct: 151 RLAV 154

55

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3473> which encodes the amino acid sequence <SEQ ID 3474>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1254-

bacterial cytoplasm --- Certainty=0.3445(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 101/127 (79%), Positives = 117/127 (91%)

Query: 1 MGGDYVLSILIDKPGGITVEDTAQLTDVVSPLLDTIQDPFPQYMLEVSSPGLERPLKT 60
 MG DY+LSIL+DK GGITVEDT+ LT+++SPLLDTI PDPFP QYMLEVSSPGLERPLKT
 10 Sbjct: 52 MGSDYILSILVDKEGGITVEDTSDLTNIISPLLDITIDPFPNQYMLEVSSPGLERPLKT 111
 Query: 61 AEALSNAVGSYINVSlyKSIDKVKIFEGDLLSFDGETLTIDYMDKTRHKTVDIPYQTVAK 120
 A++L AVGSYINVSly++IDKVK+F+GDLL+FDGETLTIDY+DKTRHK V+IPYQ VAK
 15 Sbjct: 112 ADSLKAAVGSYINVSlyQAIDKVKVFQGDLLAFDGETLTIDYLDKTRHKIVNIPYQAVAK 171
 Query: 121 ARLAVKL 127
 R+AVKL
 Sbjct: 172 VRMAVKL 178

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1123

A DNA sequence (GBSx1198) was identified in *S.agalactiae* <SEQ ID 3475> which encodes the amino acid sequence <SEQ ID 3476>. This protein is predicted to be a utilization substance protein a homolog
 25 (nusA). Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.5069(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9565> which encodes amino acid sequence <SEQ ID 9566>
 35 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13533 GB:Z99112 nusA [Bacillus subtilis]
 Identities = 164/370 (44%), Positives = 251/370 (67%), Gaps = 15/370 (4%)
 40 Query: 4 MSKEMLEAFRILEEKHINKEDIIDAVTESLKSAYKRRYQSESCVIEFNEKKADFTVYT 63
 MS E+L+A ILE+EK I+KE II+A+ +L SAYKR + Q+++ ++ N + V+
 Sbjct: 1 MSSELLDALITILEKEKGISKEIIEAIEAALISAYKRNFNQANVRVDLNRETGSIRVFA 60
 Query: 64 VREVVDDEVFDSRLAISLKDALAISAYELGDKIRFEESVTEFGRVAAQSAKQTIMEKMRR 123
 ++VVDEV+D RLEIS+++A I Y +GD + E + +FGR+AAQ+AKQ + +++R
 45 Sbjct: 61 RKDVVDDEVYDQRLAISIEEAQGIHPEYMGVDVVEIEVTPKDFGRIAAQTAQVVTQVRVE 120
 Query: 124 QMREVTTFNEYKQHEGEIMTGTVERFDQRFIYVNLGSLEAQLSHQDQIPGESFKSHDMIDV 183
 R V ++E+ E +IMTG V+R D +FIYV+LG +EA L +Q+P ES+K HD I V
 50 Sbjct: 121 AERGVIYSEFIDREEDIMTGIVQRLDNKFIYVSLGKIEALLPVNEQMPNESYKPHDRIKV 180
 Query: 184 YVYKVENNPKGVNVFVSRSHPEFIKRIMEREIPEVFDGTVEIMSVSREAGDRTKVAVRSH 243
 Y+ KVE KG ++VSR+HP +KR+ E E+PE++DGTVE+ SV+REAGDR+K++VR+
 55 Sbjct: 181 YITKVEKITKGPQIYVSRTHPGLLRLFEIEVPEIYDGTVELKSVAREAGDRSKISVRTD 240
 Query: 244 NSNVDAIGTIVGRGGSNIKKVISNFHPRVDAKTGLEIPVEENIDVIQWVEDPAEFIYNA 303
 + +VD +G+ VG G ++ +++ E ID++ W DP EF+ NA
 Sbjct: 241 DPDVPVPGSCVGPKGQRVQAIVNELK-----GEKIDIVNWSSDPVEFVANA 286

-1255-

Query: 304 IAPAEVDMVLFDDEDTKRATVVVPDSKLSLAIGRRGQNVRLAAHLTGYRIDIKSASEYEK 363
 ++P++V V+ ++E+ K TV+VPD +LSLAIG+RGQN RLAA LTG++IDIKS ++ +
 Sbjet: 287 LSPSKVLDVIVNEEE-KATTIVIPDYQLSLAIGKRGQNARLAAKL/GWKIDIKSETDARE 345

Query: 364 MEAQELQTEE 373
 + + EE
 Sbjet: 346 LGIYPRELEE 355

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3477> which encodes the amino acid sequence <SEQ ID 3478>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2074(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 348/380 (91%), Positives = 361/380 (94%), Gaps = 2/380 (0%)

- Query: 4 MSKEMLEAFRILEEKHINKEDIIDAVTESLKSAYKRRYGQSESCVIEFNEKKADFTVYT 63
 MSKEMLEAFRILEEKH+K DIIDAVTESLKSAYKRRYGQSESCVIEFNEK ADF V+T
- 25 Sbjet: 12 MSKEMLEAFRILEEKHIDKADIIDAVTESLKSAYKRRYGQSESCVIEFNEKTADFQVFT 71
- Query: 64 VREVVDEVFDSRLEISLKDALAISSAYELGDKIRFEESVTEFGRVAAQSAKQTIMEKMRR 123
 VREVV+EVFDSRLEISLKDALAISSAYELGDKIRFEESV EFGRVAAQSAKQTIMEKMRR
- 30 Sbjet: 72 VREVVVEVFDSRLEISLKDALAISSAYELGDKIRFEESVNEFGRVAAQSAKQTIMEKMRR 131
- Query: 124 QMREVTFNQYKQHEGEIMTGTVERFDQRFIYVNLGSLEAQLSHQDQIPGESFKSHDMIDV 183
 QMREV FNEYK+HEGEIMTGTVERFDQRFIYVNLGSLEAQLSHQDQIPGE+FKSHD IDV
- Sbjet: 132 QMREVMFNQYKEHEGEIMTGTVERFDQRFIYVNLGSLEAQLSHQDQIPGETFKSHDRIDV 191
- 35 Query: 184 YVYKVENNPKGVNVFVSRSHPEFIKRIMERIPEVFDGTVEIMSVSREAGDRTKVAVRSH 243
 YVYKVENNPKGVNVFVSRSHPEFIKRIME+EIPEVFDGTVEIMSVSREAGDRTKVAVRSH
- Sbjet: 192 YVYKVENNPKGVNVFVSRSHPEFIKRIMEQEIPEVFDGTVEIMSVSREAGDRTKVAVRSH 251
- 40 Query: 244 NSNVDAIGTIVGRGGSNIKKVISNFHFKRVDKGTGLEIPVEENIDVIQWVEDPAEFIYNA 303
 N NVDAIGTIVGRGGSNIKKVIS FHPKRVDKGTGLEIPVEENIDVIQWV+DPAEFIYNA
- Sbjet: 252 NPNVDAIGTIVGRGGSNIKKVISFHPKRVDKGTGLEIPVEENIDVIQWVDDPAEFIYNA 311
- Query: 304 IAPAEVDMVLFDDEDTKRATVVVPDSKLSLAIGRRGQNVRLAAHLTGYRIDIKSASEYEK 363
 IAPAEVDMVLFDDED KRATVVVPDSKLSLAIGRRGQNVRLAAHLTGYRIDIKSASEY++
- 45 Sbjet: 312 IAPAEVDMVLFDDEDLKRATVVVPDSKLSLAIGRRGQNVRLAAHLTGYRIDIKSASEYDR 371
- Query: 364 MEAQELQTEEVAQESEVISD 383
 +EA+ + A E V+ D
- 50 Sbjet: 372 LEAE--KEAATAVEEPVDD 389

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1124

- 55 A DNA sequence (GBSx1199) was identified in *S.agalactiae* <SEQ ID 3479> which encodes the amino acid sequence <SEQ ID 3480>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

- 60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2012(Affirmative) < succ>

-1256-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB13534 GB:Z99112 alternate gene name: ymxB~similar to
 hypothetical proteins [Bacillus subtilis]
 Identities = 46/92 (50%), Positives = 67/92 (72%), Gaps = 1/92 (1%)
 10 Query: 1 MAKTKKIPLRKS VVS GEVIDKRDLLRIVKNKEGQVFIDPTGKQNGRGAYIKLDNDEAILA 60
 M K KKIPLRK VV+GE+ K++L+R+V++KEG++ +DPTGK+NGRGAY+ LD + + A
 Sbjct: 1 MNKHKKIPLRKC VVTGEMKPKKELIRVRSKEGEISVDPTGKQNGRGAYLTLDKECILAA 60
 Query: 61 KKKRVFDRSFSMEVSDEFYDELLAYVDHKVKR 92
 KKK F ++ D+ +DELL + KVK+
 15 Sbjct: 61 KKKNTLQNFQSQIDDQIFDELLELAE-KVKK 91

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3481> which encodes the amino acid sequence <SEQ ID 3482>. Analysis of this protein sequence reveals the following:

20 Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1008 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 77/98 (78%), Positives = 92/98 (93%)
 30 Query: 1 MAKTKKIPLRKS VVS GEVIDKRDLLRIVKNKEGQVFIDPTGKQNGRGAYIKLDNDEAILA 60
 M+K KKIPLRKS+VS GE+I KRDLLRIVK K+GQVFIDPTGKQNGRGAYIKLDN EA++A
 Sbjct: 2 MSKVKKIPLRKSLSVSGEIIAKRDLLRIVKTKDGQVFIDPTGKQNGRGAYIKLDNQEALMA 61
 Query: 61 KKKRVFDRSFSMEVSDEFYDELLAYVDHKVKRRELGLE 98
 35 KKK+VF+RSFSM++ + FYD+L+AYVDHK+KRELGL+
 Sbjct: 62 KKKQVFNRFSMDIPESFYDDLIAVDHKIKRELGLD 99

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1125

A DNA sequence (GBSx1200) was identified in *S.agalactiae* <SEQ ID 3483> which encodes the amino acid sequence <SEQ ID 3484>. This protein is predicted to be probable ribosomal protein in infb 5'region. Analysis of this protein sequence reveals the following:

45 Possible site: 19
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:BAB06133 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 46/95 (48%), Positives = 65/95 (68%), Gaps = 1/95 (1%)
 Query: 6 KVLNLIGLAQRAGRLITGEELVIKAIQNQQVSLIFLANDAGPNLTKKVTDKSNYYKTEVS 65
 K L+L+GLA RA +L+TGEE V+KA+QN QV+L+ L++DAG + KK+ DK Y+ V

-1257-

Sbjct: 5 KWLSLILGLAARARQLLTGEEQVVKAVQNGQVTLVILSSDAGIHTKKLLDKCGSYQIPVK 64

Query: 66 TVFSTLELSDALGK-PRKVVAVADAGFSKMMRTLM 99

V + L A+GK R V+ V DAGFS+K+ L+

Sbjct: 65 VVGNRQMLGRAIGKHERVVIGVKDAGFSRKLALAI 99

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3485> which encodes the amino acid sequence <SEQ ID 3486>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1950(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 75/99 (75%), Positives = 88/99 (88%)

Query: 1 MNNSEKVLNLIGLAQRAGRLITGEELVIKAIQNQQVSLIFLANDAGPNLTKKVTDKSNYY 60

+ N E++ +LIG AQRAG++I+GEELV+KAIQ+QQV L+FLANDAGPN+TKKVTDKSNYY

Sbjct: 1 LTNLERLSSLIGPAQRAGKVISGEELVVKAIQHQVILVFLANDAGPNVTKKVTDKSNYY 60

Query: 61 KTEVSTVVFSTLELSDALGKPRKVVAVADAGFSKMMRTLM 99

EVSTV + LELS ALGKPRKV A+ADAGFSKMMRTLM

Sbjct: 61 NVEVSTVLNALELSAALGKPRKVAATADAGFSKMMRTLM 99

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1126

A DNA sequence (GBSx1201) was identified in *S.agalactiae* <SEQ ID 3487> which encodes the amino acid sequence <SEQ ID 3488>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2873(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10959> which encodes amino acid sequence <SEQ ID 10960> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3489> which encodes the amino acid sequence <SEQ ID 3490>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2985(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 735/961 (76%), Positives = 805/961 (83%), Gaps = 42/961 (4%)

-1258-

Query: 1 MSKKRLHEIAKEIGKTSKEVVEQAQSIGLPVKSHASSVEENDATRIVESFS-SSKTKAPT 59
 +SKKRLHEIAKEIGK+SKEVVE A+ LGL VKSHASSVEE DA +I+ SFS +SK
 Sbjct: 1 LSKKRLHEIAKEIGKSSKEVVEHAKYLGLDVKSHASSVEEADAKLISSFSKASKPDVTA 60

5 Query: 60 NSVQTNQGVKTESKTVETKQGLSDDKPSTQPVAKPKQSRNFKAEREARAKAEAEKRQHN 119
 + + V S TV + G S+ TQ V+KPK SRNFKAEREARAK +A ++Q N
 Sbjct: 61 SQTVPKPEVAQPSVTVVKETG-SEHVEKTQ-VSKPK--SRNFKAEREARAKEQAARKQAN 116

10 Query: 120 GD-----HRKNNRHNDTRSDRR--HQGQKRSNGNR-----NDNRQ--G 154
 G +R+ N H D+R H+ Q +N R +DN Q G
 Sbjct: 117 GSSHRSQERRGGYRQPNNHQTNEQGDKRITHRSQGDNDKRIERKASNVSPRHDNHQLVG 176

15 Query: 155 QQNN----RNKNDGRYADHKQKQTRPQQPAGNRIDFKARAAALKAEQNAEYSRHSERQRF 210
 +N N +GR+ + K++ + PQ + +IDFKARAAALKAEQNAEYSR SE RF
 Sbjct: 177 DRNRSFAKENHKNRFTNQKKQGRQEPQSKSP-KIDFKARAAALKAEQNAEYSRQSETRF 235

20 Query: 211 REEQEAKRQAAKEQELAKAAALKAEQAQKAKEKLASKPVAKVKEIVNKVAATPSQTADS 270
 R +QEAKR A ++ AK AALKAQ E +E A K + + + + TAD+
 Sbjct: 236 RAQQEAKRLAELARQEAKAALKAQAEEMSHREA-ALKSIEEAETKLKSSNISAKSTADN 294

25 Query: 271 RRKKQTRSDKSRQFSNENEDGQKQTRNKNWNQNVQVRNQRNSNWNHNKKNKKGK----T 326
 RRKKQ R +K+R+ ++ +++GQK +NKK+WN+QNQVRNQ+NSNWN NKK KKGK T
 Sbjct: 295 RRKKQARPEKNRELTHSQEGQK--KNKKSWSNQVRNQKNSNWNKNKKTCKGKNVKNT 352

30 Query: 327 NGAPKPVTERKFHELPEFEYTEGMTVAEIAKRIKREPAEIVKKLFMMGMVMTQNSLDG 386
 N APKPVTERKFHELPEFEYTEGMTVAEIAKRIKREPAEIVKKLFMMGMVMTQNSLDG
 Sbjct: 353 NTAPKPVTERKFHELPEFEYTEGMTVAEIAKRIKREPAEIVKKLFMMGMVMTQNSLDG 412

35 Query: 387 DTIELLMVDYDIEAHAKVEVDEADIERFFADEYLNPDNLTERPPVVTIMGHVDHGKTTL 446
 DTIELLMVDYDIEA AKVEVD+ADIERFF DE+YLNPD+N+ ER PVVTIMGHVDHGKTTL
 Sbjct: 413 DTIELLMVDYDIEAKAKVEVDDADIERFFEDENYLNPNENIVERAPVVTIMGHVDHGKTTL 472

40 Query: 447 LDTLRNSRVATGEAGGITQHIGAYQIEEAGKKITFLDTPGHAAFTSMRARGASVTDITIL 506
 LDTLRNSRVATGEAGGITQHIGAYQIEEAGKKITFLDTPGHAAFTSMRARGASVTDITIL
 Sbjct: 473 LDTLRNSRVATGEAGGITQHIGAYQIEEAGKKITFLDTPGHAAFTSMRARGASVTDITIL 532

45 Query: 507 IVAADDGVMPQTVEAINHNSKAAGVPIIVAINKIDKPGANPERVISELAEHGVISTAWGGE 566
 IVAADDGVMPQT+EAINHNSKAAGVPIIVAINKIDKPGANPERVI+ELAE+G+ISTAWGGE
 Sbjct: 533 IVAADDGVMPQTIEAINHNSKAAGVPIIVAINKIDKPGANPERVIAELA EYGIISTAWGGE 592

50 Query: 567 SEFVEISAKFGKNIQELLETVLLVAEMEELKADADVRAIGTVIEARLDKKGKAVATLLVQ 626
 EFVEISAKF KNI ELLETVLLVAE+EELKAD VRAIGTVIEARLDKKGGA+ATLLVQ
 Sbjct: 593 CEFVEISAKFNKNIDELLETVLLVAEVEELKADPTVRAIGTVIEARLDKKGKAIATLLVQ 652

55 Query: 627 QGTLNVQDPIVVGNTFGRVRAMNDLGRVRKVAGPSTPVSTGLNEAPMAGDHFAVYADE 686
 QGTL+VQDPIVVGNTFGRVRAM NDLGRVRK A PSTPVSTGLNE PMAGDHFAVYADE
 Sbjct: 653 QGTLHVQDPIVVGNTFGRVRAMVNDLGRVRKSAEPSTPVSTGLNETPMAGDHFAVYADE 712

60 Query: 687 KAARAAGEERAKRALLKQORQNTQRVSLNLFDTLKAGEVKSNNVVIKADVQGSVEALAAS 746
 KAARAAGEER+KRALLKQORQNTQRVSL+NLFDTLKAGE+K+VNNVVIKADVQGSVEALAAS
 Sbjct: 713 KAARAAGEERSKRALLKQORQNTQRVSLNLFDTLKAGEIKTVNNVVIKADVQGSVEALAAS 772

65 Query: 747 LLKIDVEGVKVVNVVHSAVGAINESDVTLAEASNAVIIGFNVVRPTQARQQADDDVEIRQ 806
 L+KI+VEGV+VNVVHSAVGAINESDVTLAEASNAVIIGFNVVRPTQARQQAD DDVEIR
 Sbjct: 773 LVKIEVEGVRVNVVHSAVGAINESDVTLAEASNAVIIGFNVVRPTQARQQADTDDVEIRL 832

Query: 807 HSIYKVIIEVEEAMKGLDPEYQEKILGEAIIRETFKVSKVGTIGGFMVINGKVTRDSS 866
 HSIYKVIIEVEEAMKGLDP YQEKILGEAIIRETFKVSKVGTIGGFMVINGKVTRDSS
 Sbjct: 833 HSIYKVIIEVEEAMKGLDPVYQEKILGEAIIRETFKVSKVGTIGGFMVINGKVTRDSS 892

Query: 867 VRVIRDGVVIFDGKLASLKHKKDDVKEVGNAQEGGLMIENYNDLKEDDTIEAYIMEEIKRK 927
 VRVIRD VVIFDGKLASLKHKKDDVKEVGNAQEGGLMIEN+NDLK DDTIEAYIMEEI RK
 Sbjct: 893 VRVIRDSVVIIFDGKLASLKHKKDDVKEVGNAQEGGLMIENFNDLKVDVDTIEAYIMEEIVRK 953

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1127

A DNA sequence (GBSx1202) was identified in *S.agalactiae* <SEQ ID 3491> which encodes the amino acid sequence <SEQ ID 3492>. This protein is predicted to be ribosome binding factor A (rbfA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 56
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2557(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9567> which encodes amino acid sequence <SEQ ID 9568> was also identified.

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3493> which encodes the amino acid sequence <SEQ ID 3494>. Analysis of this protein sequence reveals the following:

```

    Possible site: 60
    >>> Seems to have no N-terminal signal sequence

20    ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4765(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

25 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 93/117 (79%), Positives = 103/117 (87%)

    Query: 8   LIMANHRIDRVGMEIKREVNEILRLRVNDPRVQDVTTITDVQMLGDLSMAKVFTYTIHSTLA 67
              + MANHRIDRVGMEIKREVN+IL+ +V DPRVQ VTIT+VQM GDLS+AKV+YTI S LA
30    Sbjct: 1   MAMANHRIDRVGMEIKREVNDILQKKVRDPRVQGVTTITEVQMGGDLSLAKVYYTIMSDLA 60

    Query: 68   SDNQKAQIGLEKATGTIKRELGKNLTMYKIPDLQFVKDESYGNKIDEMLRNLDKK 124
              SDNQKAQ GLEKATGTIKRELK LTMYPKIDL F KD SI YGNKID++LR+LD K
35    Sbjct: 61   SDNQKAQTGLEKATGTIKRELKQLTMYKIPDLVFEKDNSIAYGNKIDQLLRDLDNK 117

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1128

40 A DNA sequence (GBSx1203) was identified in *S.agalactiae* <SEQ ID 3495> which encodes the amino acid sequence <SEQ ID 3496>. This protein is predicted to be esterase. Analysis of this protein sequence reveals the following:

```

    Possible site: 28
    >>> Seems to have a cleavable N-term signal seq.

45    ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

50 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA79277 GB:M64783 acetyl-hydrolase [Streptomyces hygroscopicus]
Identities = 58/220 (26%), Positives = 90/220 (40%), Gaps = 8/220 (3%)

```

-1260-

Query: 98 WNDNGKANQKTIFYLAGGSYLNPTPYHISMLKTLSTSLDAKIILPIYPKTPRYTYDYAI 157
 W + + +T+ YL GGSY H + L + A ++ Y + P + A+
 Sbjct: 58 WVRPARQDGRITLLYLHGGSYALGSPQSHRHLSSALGDAAGAAVLALHYRRPPESPFAAV 117

5 Query: 158 PRLVNLRYHFHEKN---ANLTLMGDSAGGGLALGLAHALSHQSGQEAIQPKNIILLSPW 214
 V YR E+ +TL GDSAG GLA+ AL P P + +SPW
 Sbjct: 118 EDAVAAYRMLLEQGCPPGRVTLAGDSAGAGLAVAALQALR----DAGTPLPAAAVCISPW 173

10 Query: 215 LDVTMKHPEIPKYEDTDPILSAWGLARVGEIWANGSNNTNHTYVSPKNAPATKLAPITLF 274
 D+ + + +L L R+ E + G+ + H SP + T L P+ +
 Sbjct: 174 ADLACEGASHTTRKAREILLDTADLRMAERYLAGT-DPRHPLASPAHGDLTGLPPLLIQ 232

Query: 275 TGTREIFFPDIRDYAAQLQAANHPVNYIAQEGMNHVYPIY 314
 G+ E+ D R A PV + M HV+ Y
 15 Sbjct: 233 VGSEEV LHDDARALEQAALKAGTPTVTFEEWPEMFHVHWHY 272

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3497> which encodes the amino acid sequence <SEQ ID 3498>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/334 (73%), Positives = 280/334 (83%), Gaps = 6/334 (1%)

30 Query: 1 MKPSFKKLLLLFSIITILSIACTPHAKASGRSWKSWFIEQYFWLKRDKSYKQVDESSFQ 60
 +K +K L+ ++ L + TP A AS RSWKSWFIEQYFWLKRDKSY QD+ SFQ
 Sbjct: 1 LKHPIRKTLVTLGLLLTCLP-TPVA-ASSRSWKSWFIEQYFWLKRDKSYKQDDPSFQ 58

35 Query: 61 KYLNASREQSDKGYLLDPNSVNGGLVQERLFDMQVYSWNDNGKANQKTIFYLAGGSYLN 120
 +YL+A REQSDK Y LD N VNG LVQE L+ MQVYSWNDNGK +QKTI YLAGGSYLN
 Sbjct: 59 RYLDACREQSDKPYQLDTNLVNGPLVQENLYGMQVYSWNDNGKPDQKTIYLAGGSYLN 118

Query: 121 PTPYHISMLKTLSTSLDAKIILPIYPKTPRYTYDYAIPRLVNLRYHFHEKNANLTLMGDS 180
 PT YHI+MLKTLSTSLDAKI+LPIYPK PRYTY+Y +P+LVNLY+H++ KN N+ LMGDS
 40 Sbjct: 119 PTYHINMLKTLSTSLDAKIVLPIYPKAPRYTYNYTMPKLVNLYQHYYHKNQNVFLMGDS 178

Query: 181 AGGGLALGLAHALSHQSGQEAIQPKNIILLSPWLDVTMKHPEIPKYEDTDPILSAWGLA 240
 AGGGLALGLAHAL + E++PQPK ++LLSPWLDVTM HPEIP+YED DPILS+WGL
 45 Sbjct: 179 AGGGLALGLAHALHN---ESVPQPKQLVLLSPWLDVTMSHPEIPEYEDADPILSSWGLK 234

Query: 241 RVGEIWANGSNNTNHTYVSPKNAPATKLAPITLFTGTREIFFPDIRDYAAQLQAANHPVN 300
 RVGE+WA ++NTNH YVSPKN P T L PITLFTGTREIF+PDIRDYAA+L+AANH +
 Sbjct: 235 RVGELWAYSADNINHIYVSPKNGPITLPPITLFTGTREIFPDIRDYAAKLKAANHNT 294

50 Query: 301 YIAQEGMNHVYPIYPIEEAKTAQYQIMIDIINKTP 334
 +I QEGMNHVYPIYPIEEAKTAQYQ+ID INKTP
 Sbjct: 295 FITQEGMNHVYPIYPIEEAKTAQYQIIDAINKTP 328

A related GBS gene <SEQ ID 8731> and protein <SEQ ID 8732> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6

McG: Discrim Score: 11.88

GvH: Signal Score (-7.5): -1.33

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 0 value: 4.03 threshold: 0.0

PERIPHERAL Likelihood = 4.03 174

modified ALOM score: -1.31

*** Reasoning Step: 3

```
5      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

```

10      28.4/46.2% over 220aa
                                                    Streptomyces
hygroscopicus
    EGAD|5925| acetyl-hydrolase Insert characterized

15      ORF00486(589 - 1245 of 1602)
    EGAD|5925|5724(57 - 277 of 300) acetyl-hydrolase {Streptomyces hygroscopicus}
    %Match = 6.8
    %Identity = 28.3  %Similarity = 46.1
    Matches = 62  Mismatches = 111  Conservative Sub.s = 39

20      462          492          522          552          582          612          642          669
    KRDKSYKVKQDESSFOKYLNASREQSDKGYYLDPNSSVNGGLVQERLFDMQVYSWNDNGKANQKTIIFYLAGGSY-LNNPTP
      :      ::      :      |      :      ::      |      :      :      :|||  | :
    ELELVRELIELNWHTRNGEMEPRIIAYDRAQEAFGNLGVPPGDVVTVGHCTAEWVRPARQDGRLLLYLHGGSYALGSPQS

25      20          30          40          50          60          70          80
    696          726          756          786          837          867          897
    Y-HISMLKTLSTSLDAKIILPIYPKTPRYTYDYAIPRLVNLVYRHFHEKN---ANLTLMGDSAGGGLALGLAHALSHQSGQ
      :||  | :  | :  | :  | :  | :  | :  | :  | :  | :  | :  | :  | :  | :  | :  | :
    HRHLS--SALGDAAGAAVLALHYRRPPESPFPAAVEDAVAAYRMLLEQGCPPGRVTLAGDSAGAGLAVAALQAL----RD

30      100          110          120          130          140          150
    927          957          987          1017          1047          1077          1107          1137
    EAIPQPKNIILLSPWLDVTMKHPEIPKYEDTDPILSANGLARVGEIANGSNNNTNHTYVSPKNAPATKLAPITLFTGTRE
      ||  : :||| | :  :  : :|  | :| : | :| :|  || :  | | : : | : |
    AGTPLPAAAVCI SPWADLACEGASHTTRKAREILLDTADLRMAERYLAGTD-PRHPLASPAHGDLTGLPPLLIQVGSEE

35      170          180          190          200          210          220          230
    1167          1197          1227          1245          1275          1305          1335          1365
    IFFPDIRDYAAQLQAANHPVNYIAQEGMNHV----YPIYPIEEAKTAQYQMIDIINKTP*Y*LSQL*SYKK*TMILTWFI
      ::  ||  | || :  ||  :||  :  :  :  :  :  :  :  :
    VLHDDARALEQAALKAGTVPVTFEEWPEMFHVWHVHPVLPEGRRAAIEVAGAFRLRTATGEGLK

40      250          260          270          280          290          300

```

The GBS149-His fusion product was purified (Figure 196, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 291), which confirmed that the protein is immunoaccessible on GBS bacteria.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A DNA sequence (GBSx1204) was identified in *S.agalactiae* <SEQ ID 3499> which encodes the amino acid sequence <SEQ ID 3500>. This protein is predicted to be CopY. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```


-1262-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3140(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG10085 GB:AF296446 CopY [Streptococcus mutans]
 Identities = 67/137 (48%), Positives = 98/137 (70%)

10 Query: 2 TISSAEWEIMRVVWAQQNTTSNEILAVLLEKYDWPSTVKTLRLRLDQGYVSREKMGKG 61
 +IS+AEWE+MRVVWA+Q T+S+EI+A+L Y W+ ST+KTL+ RL +KGY++ ++ G+
 Sbjct: 3 SISNAEWEVMRVVWAKQMTSSSEIIILSRITYCWSASTIKTLITRLSEKGYLTSQRQGRK 62

15 Query: 62 FSYSPLIDEDLAMMSEVDSVFQKVCQTKHVAIVRHLESIPMTEKDRLNLQSSLEAKKGK 121
 + YS LI E+ A+ +V VF ++C TKH A++RHL+E PMT D L++ L +KK
 Sbjct: 63 YIYSSLISEEEALEQQVSEVFSRICVTKHQALIRHLVEETPMTLSDIEKLEALLSKKAN 122

20 Query: 122 TLERVACNCIPGQCQCH 138
 + V CNCI GQC C+
 Sbjct: 123 AVPEVKCNCIVGQCSCY 139

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3501> which encodes the amino acid sequence <SEQ ID 3502>. Analysis of this protein sequence reveals the following:

25 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2331(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 54/135 (40%), Positives = 84/135 (62%)

35 Query: 3 ISSAEWEIMRVVWAQQNTTSNEILAVLLEKYDWPSTVKTLRLRLDQGYVSREKMGKGF 62
 IS+AEWE+MRVVWA + S++I+ +L +KY W+ ST+KTL+ RL+ K +++ + G+ +
 Sbjct: 10 ISAAEWEVMRVVWASGDIKSSDIITILRKKYQWSDSTIKTLIGRLVKKNFLTYSYRQGRAY 69

40 Query: 63 SYSPLIDEDLAMMSEVDSVFQKVCQTKHVAIVRHLESIPMTEKDRLNLQSSLEAKKGKT 122
 Y L+DE L + +V +CQ +H ++ L +PMT ++ Q LE KK
 Sbjct: 70 IYQALLDETLLQKEALATVLDGICQQRQHTRLLLERLYHLPMTLEEIGAFQELLEVKKENA 129

45 Query: 123 LERVACNCIPGQCQC 137
 + V CNC+PGQC C
 Sbjct: 130 VLEVPCNCLPGQCHC 144

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1130

A DNA sequence (GBSx1206) was identified in *S.agalactiae* <SEQ ID 3503> which encodes the amino acid sequence <SEQ ID 3504>. This protein is predicted to be CopA. Analysis of this protein sequence reveals the following:

55 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.82	Transmembrane	382 - 398 (370 - 406)
INTEGRAL	Likelihood = -8.01	Transmembrane	356 - 372 (344 - 374)
INTEGRAL	Likelihood = -2.50	Transmembrane	719 - 735 (719 - 738)

-1263-

INTEGRAL Likelihood = -2.28 Transmembrane 202 - 218 (202 - 218)
 INTEGRAL Likelihood = -1.59 Transmembrane 693 - 709 (691 - 712)
 INTEGRAL Likelihood = -1.33 Transmembrane 167 - 183 (167 - 183)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG10086 GB:AF296446 CopA [Streptococcus mutans]
 Identities = 440/740 (59%), Positives = 571/740 (76%), Gaps = 1/740 (0%)

15 Query: 5 KETFLIDGMTCASCALTIEKAVNKLDHVD SAVVNLATEKMTVTFFDDTTLSPNVIEECVSE 64
 +E FLIDGMTCASCA+ +E AV KLD ++SAVVNL TEKMT+ +D +S + + V+
 Sbjct: 3 EEVFLIDGMTCASCAINVENAVKKLDGIESAVVNLTEKMTIDYDAAKVSEADVTKAVAG 62

20 Query: 65 SGYEASLFKEETSKSQSERHQIAIEKMWHRFWMSAVATIPLLYISMGPMINLWLP SFLMP 124
 +GY A ++ T++SQ +R + + + R +++ TIPL YI+MG M+ L LP+FL P
 Sbjct: 63 AGYGAKVYDPTTAESQKDR EHKLAGIKRLLWTSIFTIPLFYIAMGSMVGLPLPNFLAP 122

25 Query: 125 DKGPLNYGMIQ LLLTL PVMYFGRIFYQNGFKALFKRHPNMDSLVAIATTA AFIYSLYGLY 184
 PL Y M+ LLLT+PV+ FY NGF++LFK HPNMDSLV++ATTA AF+YSLYG Y
 Sbjct: 123 SSAPLTYAMV LLLLTIPVIVLSWSFYDNGFRSLFKGHPNMDSLVSLATTA AFLYSLYGTY 182

30 Query: 185 EILQGDIIHYAHQLYFESVAVILTLITLGKYFEILSKGRTSASIEKLLT LSAKEARVIKDG 244
 + G H+AH LY+ESVAVILTLITLGKYFE LSKGRTS +I+KL+ LSAKEA +I+DG
 Sbjct: 183 HVYLGHTTHAHHLHYESVAVILTLITLGKYFETLSKGRTSDAIKKLMHLSAKEATLIRDG 242

35 Query: 245 EDYMVPLDKVKIGETILVKG EKIPLDGHV VAGESSIDESMLTGESIPVEKKVGSKVYGA 304
 E+ VP+++V+I + ILVKPGEKIP+DG V++G S+IDESMLTGESIP+EK S VY
 Sbjct: 243 EEIKVPBQVQIRDQILVKPGEKIPVDGRVLSGHS AIDESMLTGESIPIEKMADSPVYAG 302

40 Query: 305 SINGQGS LTI FVEKEAGG SLLSQI IINLVEAAQTSKAPIANLADK VSGVFVPFVIVIAILS 364
 SINGQGS LTI EK +LLSQII LVE AQ +KAPIA +ADKVS VFVP +I IAIL+
 Sbjct: 303 SINGQGS LTI FEA EKVGNETLLSQI IKLVENAQQTKAPIAKIADK VSAVFVPIITAILT 362

45 Query: 365 GLSWYLILGQSFAFSLKIMIAVLVIACPCALGLATPTAIMVASGKAAENGILFKGGEVLE 424
 GL WY ++GQ F FS+ I +AVLVIACPCALGLATPTAIMV +G+AAENGIL+K G+VLE
 Sbjct: 363 GLFWYFVMGQDFTFSMTISAVLVIACPCALGLATPTAIMVGTGRAAENGILYKRGDVLE 422

50 Query: 425 KAHHIDTIVFDKTGTITLTKGKPEVVAIKTYGGDK EEFLGQVASVEKLSNHPLSQTIVNKAK 484
 AH I+TIVFDKTGT+T+GKPEV +Y D+ + + A++E LS HPLSQ IV+ AK
 Sbjct: 423 LAHQINTIVFDKTGTITQKGPEVVHQFSY-HDRTDLVQVTAAL EALSEHPLSQAIVDYAK 481

55 Query: 485 EKELPLREVMAFKNILGYGLSATINKTMLVGNANLMTKNDVNLDLAKADIEIAQEEAQT 544
 ++ L V F ++ G GL + +T+LVGN LM + +++L+ A+AD + A + QT
 Sbjct: 482 KEGTHLLAVDDFTSLTGLGLKGCVADETLLVGN EKL MRQANISLEQAQADFKAAATAQGQT 541

60 Query: 545 VVYVSENGVLSGLITLTDQLKTD SQETVKQLQRLGFNLVLLTGDNKASADAIAQKLGITT 604
 ++V+ +G L GLIT+ D++K DS TVK LQ +G + +LTGDN+ +A AIA+++GIT
 Sbjct: 542 PIFVASDQQLGLIT IADKVKND SAATVKALQNMGVEVAMLTGDNEETAQAIAKEVGITF 601

65 Query: 605 VVSEVLPDQKANVILELKEKGGQIAMVGDGINDAPALASSDVGISMS SGT DIAIESADIV 664
 V+S+V +K IL+L+ +G ++AMVGDGINDAPALA++D+GISM SGT DIA+ESADIV
 Sbjct: 602 VISQVFSQ EKTQA ILDLQAEGKKVAMVGDGINDAPALATADIGISM SGT DIAMESADIV 661

Query: 665 LMKPELTDLLKAMTISKQTIQI IKENLFWAFFYNVLAIPVAMGV LHLFGGPLLPMLAGL 724
 LMKP + D++KA+ IS+ TI IKENLFWAF YNVL++P+AMGV L+LFGGPLLP+PM+AGL
 Sbjct: 662 LMKPAMLDI I KALKISRVTIINIKENLFWAFIYNVLSVP IAMGVLYLFGGPLLPDPMIAGL 721

Query: 725 AMAFSSSVSVVLNALRLKVLK 744
 AM+FSSSVSVVLNALRLKV+K
 Sbjct: 722 AMSFSSSVSVVLNALRLKVVK 741

There is also homology to SEQ ID 3506.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1131

A DNA sequence (GBSx1207) was identified in *S.agalactiae* <SEQ ID 3507> which encodes the amino acid sequence <SEQ ID 3508>. This protein is predicted to be cation-transporting ATPase, P-type (pacS). Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1934 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG10087 GB:AF296446 CopZ [Streptococcus mutans]
Identities = 31/67 (46%), Positives = 43/67 (63%)

Query: 1 MKHTYRVSGMKCDGCAKTVSDKLSVIGVDEVNVDLTKNQVVVSGKTFKWLLKRSLKDTK 60
M+ TY + G+KC GCA V+ + S + V++V VDL K +V ++G KW LKR+LK T
Sbjct: 1 MEKTYHIDGLKCGCADNVTKRFSELKKVNDVKVLDLKKVIRITGNPSKWSLKRALKGTN 60

Query: 61 YSLEEEI 67
Y L EI

Sbjct: 61 YELGAEI 67

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3509> which encodes the amino acid sequence <SEQ ID 3510>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2997 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 33/63 (52%), Positives = 48/63 (75%)

Query: 1 MKHTYRVSGMKCDGCAKTVSDKLSVIGVDEVNVDLTKNQVVVSGKTFKWLLKRSLKDTK 60
M+ Y+V+GM CDGCA+TV++KLS+V GV V V+L K + V+G+ +L+KR+LKDTK
Sbjct: 1 MEKHYQVTGMTCDGCARTVTEKLSAVPGVQSVQVNLEKGEAKVTGRPLTFLLIKRALKDTK 60

Query: 61 YSL 63

+ L

Sbjct: 61 FEL 63

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1132

A DNA sequence (GBSx1208) was identified in *S.agalactiae* <SEQ ID 3511> which encodes the amino acid sequence <SEQ ID 3512>. Analysis of this protein sequence reveals the following:

Possible site: 20

-1265-

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -7.59	Transmembrane	67 - 83 (65 - 90)
INTEGRAL	Likelihood = -3.72	Transmembrane	35 - 51 (31 - 51)
INTEGRAL	Likelihood = -3.61	Transmembrane	122 - 138 (120 - 139)
INTEGRAL	Likelihood = -1.59	Transmembrane	154 - 170 (154 - 171)

----- Final Results -----

bacterial membrane	---	Certainty=0.4036(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 8733> which encodes amino acid sequence <SEQ ID 8734> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 5

McG: Discrim Score: 4.09

GvH: Signal Score (~7.5): 3.87

Possible site: 20

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 4 value: -7.59 threshold: 0.0

INTEGRAL	Likelihood = -7.59	Transmembrane	65 - 81 (63 - 88)
INTEGRAL	Likelihood = -3.72	Transmembrane	33 - 49 (29 - 49)
INTEGRAL	Likelihood = -3.61	Transmembrane	120 - 136 (118 - 137)
INTEGRAL	Likelihood = -1.59	Transmembrane	152 - 168 (152 - 169)
PERIPHERAL	Likelihood = 0.85		96

modified ALOM score: 2.02

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane	---	Certainty=0.4036(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15351 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
Identities = 107/192 (55%), Positives = 137/192 (70%)

Query: 8 WNILSLVGTVAFASSGAIVAIEEEFDILGLFILGFVTAFGGGAIRNVLIGLPIETLWSQG 67
W +LS++G +AFA SGAIVA+EEE+DILG++ILG VTAFGGGAIRN+LIG+P+ LW QG

Sbjct: 3 WELLSVIGIIAFAVSGAIVAMEEEYDILGVYILGIVTAFGGGAIRNLLIGVPVSALWEQG 62

Query: 68 IAFYAAAAAILFIMIFPNLLSGKGRDAEVVSDAIGLAAFSVQGALYATQSHQPLSAVIVA 127
F A +I ++FP LL +SDAIGLAAF++QGALYA + PLSAVIVA

Sbjct: 63 AYFQIALLSITIVFLFPKLLKHWNKGNLSDAIGLAFAIQGALYAVKMGHPLSAVIVA 122

Query: 128 AVLTGAGGGIIVRDVLAGRKPGVLRSEIYAGWSILVGIILYFKIAKTTTDYLLVVLVVTSL 187
AVLTG+GGGI+RD+LAGRKP VL++EIYA W+ L G+I+ + Y+L V+

Sbjct: 123 AVLTGSGGGIIRDLLAGRKPLVLKAEIYAVWAALGGLIVGLGWLNSFGLYVLFVFLVVC 182

Query: 188 RMLGYKKQWHL P 199

R+ Y W LP

Sbjct: 183 RVCSYMFNWKLP 194

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3513> which encodes the amino acid sequence <SEQ ID 3514>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -5.15	Transmembrane	70 - 86 (65 - 88)
INTEGRAL	Likelihood = -4.09	Transmembrane	33 - 49 (29 - 49)
INTEGRAL	Likelihood = -2.13	Transmembrane	120 - 136 (119 - 137)
INTEGRAL	Likelihood = -0.43	Transmembrane	173 - 189 (172 - 189)

----- Final Results -----

-1266-

bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:BAB05428 GB:AP001512 unknown conserved protein [Bacillus halodurans]
 Identities = 109/195 (55%), Positives = 137/195 (69%)

10 Query: 6 WEILNIIGTIAFALSGAIVAMEEEFDILGIFILGFVTAFGGGAIRNTLIGLPIEALWGQK 65
 W++LN+IGTIAFALSG IVAMEE+FD++G++ILGFVTAFGGGAIRN LIG+P+ ALW Q
 Sbjct: 3 WDVLNVIGTIAFALSGVIVAMEEEDFDLMGVYILGFVTAFGGGAIRNLLIGVPVSALWEQG 62

15 Query: 66 PEFTCAFFAMVLIMLFPKLMARGWVRAAVLTDAGLAAFSVQGALHAVRLNQPLSAVIVT 125
 FT AF M + P L W++ +L DAIGLAAF++QGAL A ++ PLSAVIV
 Sbjct: 63 TLFTIAFIVMTIAFFLPNLWINHWLKFGLLFDAGLAAFAIQGALFATSMDHPLSAVIVA 122

20 Query: 126 AVLTGAGGGVVRDILAGRKPSVLRSEIYAGWSILAAIVLHFKLADSTIECYALVLLTTL 185
 A LTGAGGG+VRD+LA RKP VL EY GW++LA + + I , L++L+ L
 Sbjct: 123 AALTGAGGGIVRDMRLARRKPLVLSKEIYIGWAMLAGAAIGLNIVSGPIGIGFLIILVVFL 182

Query: 186 RMIGNRKKWNLPKIK 200
 RM+ W LP K
 Sbjct: 183 RMLSVHYNWCLPHRK 197

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 133/200 (66%), Positives = 168/200 (83%)

30 Query: 3 MSIDIWNILSLVGTVAFASSGAIVAIEEEFDILGLFILGFVTAFGGGAIRNVLIGLPIET 62
 M+ID+W IL+++GT+AFA SGAIVA+EEEFDILG+FILGFVTAFGGGAIRN LIGLPIE
 Sbjct: 1 MTIDMWEILNIIGTIAFALSGAIVAMEEEFDILGIFILGFVTAFGGGAIRNTLIGLPIEA 60

35 Query: 63 LWSQGIAFYAAAAAILFIMIFPNLSSGKGRDAEVVSDAIGLAAFSVQGALYATQSHQPLS 122
 LW Q F A A++ IM+FP L++ A V++DAIGLAAFSVQGAL+A + +QPLS
 Sbjct: 61 LWGQKPEFTCAFFAMVLIMLFPKLMARGWVRAAVLTDAGLAAFSVQGALHAVRLNQPLS 120

40 Query: 123 AVIVA AVL TGAGGGIVRDVLAGRKPGVLRSEIYAGWSILVGIILYFKIAKTTTDDYLLVL 182
 AVIV AVL TGAGGG+VRD+LAGRKP VLRSEIYAGWSIL I+L+FK+A +T + Y LV+
 Sbjct: 121 AVIVTAVLTGAGGGVVRDILAGRKPSVLRSEIYAGWSILAAIVLHFKLADSTIECYALVV 180

Query: 183 VVTSLRMLGYKKQWHLPPVR 202
 ++T+LRM+G +K+W+LP ++
 Sbjct: 181 LLTTLRMIGNRKKWNLPKIK 200

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 45 vaccines or diagnostics.

Example 1133

A DNA sequence (GBSx1209) was identified in *S.agalactiae* <SEQ ID 3515> which encodes the amino acid sequence <SEQ ID 3516>. Analysis of this protein sequence reveals the following:

50 Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.2805(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9569> which encodes amino acid sequence <SEQ ID 9570> was also identified.

-1267-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB94816 GB:AJ245582 hypothetical protein [Streptococcus thermophilus]
Identities = 138/238 (57%), Positives = 184/238 (76%)

```

5   Query: 5   KMKIKLIAIDMDGTLNDEKKIPKENIQAIKEATQAGIKIVLCTGRPMGILPYFNEGL 64
      +   +KLIAIDMDGTLN +K+IPKENI+AI+EAT AGIKIVLCTGRP SGI+P+F +LGL
      Sbjct: 3   QNQVKLIAIDMDGTLNLSQKEIPKENIKAIQEATAAGIKIVLCTGRPRSGIVPHFEKLG 62

10  Query: 65   TKEEYIIMNNGCSTYSTKDWQLIDSATLTHDELIFLEEVVKEFPNVCLTLTAENTFYAVG 124
      ++EE+IIMNNGCSTY TK+W L++S +L+ E+ L + ++FP V LT T E ++Y VG
      Sbjct: 63   SEEEFIIMNNGCSTYETKNWTLLESESLSRSEMEELLQACEDFPGVALTFTGEKSYVVG 122

      Query: 125  EEVPEIVAYDADLVFTKAKSTSLDALRNQEEIVFQAMYMGLDADVTAFQEAEEALISKF 184
      EVPE+VAYDA VFT+AK+ SL+ + + +++FQAMYM + AFQ AV++ L +
15  Sbjct: 123  NEVPPELVAYDAGTVFTEAKARSLKEEIFEEGQVIFQAMYMAESEPLDAFQNAVQDRLDQSY 182

      Query: 185  SGVRSQDYIYEIMPQGVTKARGLKSLIAKGLDINQVMAIGDAPNDIELLDLPNSVA 242
      S VRSQ+YI+E+MPQG TKA GLK L KL ++ +Q+MA+GDA ND+E+L V SVA
20  Sbjct: 183  STVRSQYEIFEVMPOGATKASGLKHLAEKLDINRDQIMALGDAANDEMLQFVGQSV 240

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3517> which encodes the amino acid sequence <SEQ ID 3518>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1468(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 152/270 (56%), Positives = 193/270 (71%)

```

35  Query: 6   KMKIKLIAIDMDGTLNDEKKIPKENIQAIKEATQAGIKIVLCTGRPMGILPYFNEGLT 65
      +MI+LIAID+DGTLLN +K+IPKENI AI+EA Q+G+KIVLCTGRP SG PYF++LGLT
      Sbjct: 19  RMIQLIAIDLGDGTLNLDKQIPKENITAIQEAAQSGLKIVLCTGRPQSGTRPYFDQLGLT 78

      Query: 66   KEEYIIMNNGCSTYSTKDWQLIDSATLTHDELIFLEEVVKEFPNVCLTLTAENTFYAVGE 125
      +EE++I+NNGCSTYS+ DWQL S L ++ LEE+ + FP++ LTLT EN + + E
40  Sbjct: 79   QEEFLIIMNNGCSTYSSPDWQLRHSKMLKVSDIELLEELSQSFPDIYLTTEENDYLVLEE 138

      Query: 126  EVPEIVAYDADLVFTKAKSTSLDALRNQEEIVFQAMYMGLDADVTAFQEAEEALISKFS 185
      EVP++V D DLVFT K SL L + ++FQAMY+G A + AF+ AV L F
45  Sbjct: 139  EVPDLVQEDGLVFTIVKPVSLAELSDTFLIFQAMYLGEKAALDAFERAVRNQLSQSFH 198

      Query: 186  GVRSQDYIYEIMPQGVTKARGLKSLIAKGLDINQVMAIGDAPNDIELLDLPNSVAMGN 245
      VRSQD I EI+PQGV+KA LK L+ LGL +QVMAIGDAPNDIE+L VAM N
50  Sbjct: 199  VVRSQDNILEILPQGVSKASALKELVEDLGLTADQVMAIGDAPNDIEMLTAYAGLGVAMEN 258

      Query: 246  ASDEIKSRCKYITVDNNKAGVAKAIYDYAL 275
      AS IK +T+ N+ AGVA+AI +AL
      Sbjct: 259  ASAAIKPLADKVTLTNDMAGVAQAIRQFAL 288

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1134

A DNA sequence (GBSx1210) was identified in *S.agalactiae* <SEQ ID 3519> which encodes the amino acid sequence <SEQ ID 3520>. Analysis of this protein sequence reveals the following:

Possible site: 18

-1268-

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.43 Transmembrane 7 - 23 (7 - 23)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAA26954 GB:J04479 DNA polymerase I [Streptococcus pneumoniae]
 Identities = 655/879 (74%), Positives = 748/879 (84%), Gaps = 4/879 (0%)

Query: 3 NKNKLLLLIDGSSVAFRAFFALYNQIDRFKNNSGLHTNAIYGFHMLNLHILGRVQPSHILV 62
 +K KLLLLIDGSSVAFRAFFALY Q+DRFKN +GLHTNAIYGF LML+H+L RV+PSHILV

15 Sbjct: 2 DKKKLLLLIDGSSVAFRAFFALYQQLDRFKNAAGLHTNAIYGFQMLSHLLERVEPSHILV 61

Query: 63 AFDAGKTTFRTEMYADYKGGRAKTPDEFREQFPYIRQQLDVLGIKHYELEHYEADDIIGT 122
 AFDAGKTTFRTEMYADYKGGRAKTPDEFREQFP+IR+ LD +GI+HYEL YEADDIIGT

20 Sbjct: 62 AFDAGKTTFRTEMYADYKGGRAKTPDEFREQFPFIRELLDHMGIRHYELAQYEADDIIGT 121

Query: 123 LAKQAEASNEHFDITVVGDKDLIQLTDTNTVVEISKKGVAEFEEFTPAYLMEKMGITPS 182
 L K AE + FDIT+VSGDKDLIQLTD +TVVEISKKGVAEFE FTP YLME+MG+TP+

Sbjct: 122 LDKLAE--QDGFDTITVVGDKDLIQLTDEHTVVEISKKGVAEFEAFTPDYLMEEMGLTPA 179

25 Query: 183 QFIDLKALMGDKSDNIPGVTKIGEKTGLKLLSEYGSLEGIYENIEAMKQSKMKENLINDK 242
 QFIDLKALMGDKSDNIPGVTK+GEKTG+KLL E+GSLEGIYENI+ MK SKMKENLINDK

Sbjct: 180 QFIDLKALMGDKSDNIPGVTKVGEKTGIKLLLEHGSLEGIYENIDGMKTSKMKENLINDK 239

Query: 243 EQAFLSKTLATINIASPITIGLEDILYSGPQDIKALSQFYDEMDFKQFKAALGEETSQED 302
 EQAFLSKTLATI+ +PI IGLED++YSGF D++ L +FYDEM FKQ K AL ++

30 Sbjct: 240 EQAFLSKTLATIDTKAPIAIGLEDLVYSGP-DVENLGKFYDEMDFKQKQALNMSSADVA 298

Query: 303 FEVDFTVEVEQLKTEMFSDNDFYFEMLGDNYHVEDLIGIAWNSDTIYATSNVSLLEQAL 362
 +DFT V+Q+ +M S+ ++FE+ G+NYH ++L+G AW D +YAT + LLQ+ +

35 Sbjct: 299 EGLDFTIVDQISQDMLSEBSIFHFELFGENYHTDNLVGFPAWSCGDQLYATDKLELLQDPI 358

Query: 363 FKKALSKP-IKTYDFKRSKVLNRFNIDLPAPDFTRLAKYLLSTTEDNLVSTIARLYTN 421
 FK L K ++ YDFK+ KVLL RF +DL PAFD RLAKYLLST EDN ++TIA LY

40 Sbjct: 359 FKDFLEKTSLRVYDFKKVKVLLQRFVVDLQAPAFDIRLAKYLLSTVEDNEIATIASLYGQ 418

Query: 422 LPLDTDDAVYGKAKRAIPEKTRFLEHLAKKVKVLVDSEANIMQQLKANEQEELLFEMEQ 481
 L D+ YGKG K+AIPE+ +FLEHLA K+ VLV++E ++++L N Q ELL++MEQ

Sbjct: 419 TYLVDETFYGGKGVKAIPEREKFLLEHLACKLAVLVETEPILLEKLSENGQLELLYDMEQ 478

45 Query: 482 PLANVLAKMEIRGIKVKKNLTNEMAIENQKVIETLTQEIYELAGQEFNINSKPKQLGLLF 541
 PLA VLAKMEI GI VKK TL EM EN+ VIE LTQEIYELAG+EFN+NSPKQLG LLF

Sbjct: 479 PLAFVLAKMEIAGIVVKKETLLEMQAENELVIEKLTQEIYELAGEEFNVNSPKQLGVLLF 538

Query: 542 ETLGLPVEMTKTKTGYSTAVDVLERLAPISPLVTKILEYRQITKLQSTYIIGLQDYILE 601
 E LGLP+E TKTKTGYSTAVDVLERLAPI+P+V KIL+YRQI K+QSTY+IGLQD+IL

50 Sbjct: 539 EKLGLPLEYTKTKTGYSTAVDVLERLAPIAPIVKKILDYRQIAKIQSTYVIGLQDWILA 598

Query: 602 DGKIHTRYVQDLTQTGRLLSSSDPNLQNIPIVRLEQGRLIRKAFVPSEDNAVLLSSDYSQIE 661
 DGKIHTRYVQDLTQTGRLLS DPNLQNIPI RLEQGRLIRKAFVP +++VLLSSDYSQIE

55 Sbjct: 599 DGKIHTRYVQDLTQTGRLLSSVDPNLQNIIPARLEQGRLIRKAFVPEWEDSVLLSSDYSQIE 658

Query: 662 LRVLAHISKDEHLIAAFKEGADIHTSTAMRVFGIEKPENVTPNDRRNAKAVNFGIVYGIS 721
 LRVLAHISKDEHLI AF+EGADIHTSTAMRVFGIE+P+NVT NDRRNAKAVNFG+VYGIS

60 Sbjct: 659 LRVLAHISKDEHLIKAFQEGADIHTSTAMRVFGIERPDNVTANDRRNAKAVNFGVYGIS 718

Query: 722 DFGLSHNLGIPRKLAKQYIDTYFERYPGIKNYMETVVREAKDKGYVETLFHRRRSLPDIN 781
 DFGLS+NLGI RK AK YIDTYFER+PGIKNYM+ VVREA+DKGYVETLF RRR LPDIN

Sbjct: 719 DFGLSNNLGISRKEAKAYIDTYFERFPGIKNYMDEVVREARDKGYVETLFKRRELDPIN 778

65 Query: 782 SRNFNIRQFAERTAINSPIQGSAAIDILKIAMINLDRVLDKGGYKSKMLLQVHDEIVLEVP 841
 SRNFNIR FAE TAINSPIQGSAAIDILKIAM LD+ L GGY++KMLLQVHDEIVLEVP

Sbjct: 779 SRNFNIRGFAEATAINSPIQGSAAIDILKIAMIQLDKALVAGGYQTKMLLQVHDEIVLEVP 838

-1269-

Query: 842 NEEIGAIRELVTKTMEISLVSPLIADENAGETWYEAK 880
 E+ +++LV +TME AI LSVPLIADEN G TWYEAK
 Sbjct: 839 KSELVEMKKLVKQTMEEAIQLSVPLIADENEGATWYEAK 877

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3521> which encodes the amino acid sequence <SEQ ID 3522>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.43 Transmembrane 7 - 23 (7 - 23)

----- Final Results -----

bacterial membrane --- Certainty=0.1171(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 665/881 (75%), Positives = 761/881 (85%), Gaps = 2/881 (0%)

Query: 1 MTNKNKLLIDGSSVAFRAFFALYNQIDRFKNNSGLHTNAIYGFHMLNLHILGRVQPSHI 60
 M NKNKLLIDGSSVAFRAFFALYNQIDRFKN+SGLHTNAIYGFHML+H++ RVQP+H+
 Sbjct: 1 MENKNKLLIDGSSVAFRAFFALYNQIDRFKNHSGLHTNAIYGFHMLDHHMKRVQPTHV 60

Query: 61 LVAFDAGKTTFRTEMYADYKGRAKTPDEFREQFPYIRQLDVLGKIKHYELEHYEADDII 120
 LVAFDAGKTTFRTEMYADYK GRAKTP+EFREQFPYIR+ L LGI +YELEHYEADDII
 Sbjct: 61 LVAFDAGKTTFRTEMYADYKAGRAKTPPEEFREQFPYIREMLTALGIAYYELEHYEADDII 120

Query: 121 GTLAKQAEASNEHFDITVVGDKDLIQLTDTNTVVEISKKGVAEFEFTPAYLMEKMGIT 180
 GTL K AE + FD+T+VSGDKDLIQLTD NTVVEISKKGVAEFEFTPAYLMEKMG+T
 Sbjct: 121 GTLDKMAERTEVPFDVTIVVGDKDLIQLTDENTVVEISKKGVAEFEFTPAYLMEKMGLT 180

Query: 181 PSQFIDLKALMGDKSDNIPGVTKIGEKTGLKLLSEYGSLEGIYENIEAMKQSKMKENLIN 240
 P+QFIDLKALMGDKSDNIPGVTKIGEKTGLKLL E+GSLEGIYE+I+ K SKMKENLIN
 Sbjct: 181 PNQFIDLKALMGDKSDNIPGVTKIGEKTGLKLLHEFGSLEGIYEHIDGFKTSKMKENLIN 240

Query: 241 DKEQAFLSKTLATINIASPITIGLEDILYSGPDIKALSQFYDEMDFKQFKAALGEETSQ 300
 D++QAFLSKTLATIN ASPITIGL+DI+Y+GP D+ +LSQFYDEMDF Q K L + Q
 Sbjct: 241 DRDQAFLSKTLATINTASPITIGLDDIVYNGP-DVASLSQFYDEMDFVQLKKGLASQMPQ 299

Query: 301 EDFEV-DFTEVEQLKTEMFSNDNFYFEMLGDNYHVEDLIGIAWGNSDTIYATSNVSLQ 359
 E V + EV + ++FS D +YFE L DNYH E +IG AWG+ + IYA++++ LL
 Sbjct: 300 EPVAVISQEVITNVSADLFSADIFYFETLRDNYHREAIIGFAWGHGEQIYASTDLGLLA 359

Query: 360 EALFKKALKSPIKTYDFKRSKVLNRFNIDLPEPAFDTRLAKYLLSTTEDNLVSTIARLY 419
 FK+ KPI TYDFKRSKVLN+ I+L P++D RLA YLLST EDN +STIAR++
 Sbjct: 360 TDSFKQVFQKPIATYDFKRSKVLNLSHLGIELVAPSYDARLANVLLSTVEDNELSTIARIF 419

Query: 420 TNLPLDTDDAVYGKAKRAIPEKTRFLEHLAKVKVLDSEANIMQQLKANEQEELLFEM 479
 T++ L+ DD VYGKAKRA+P+K LEHLA+KVKVL+DS++ ++ +L A+EQ +L +
 Sbjct: 420 TDISLEEDTVYGKAKRAVPDKDVLLEHLARKVKVLLDSKSQMLDKLTAHEQLDLYQNI 479

Query: 480 EQPLANVLAKMEIRGIKVKNTLNEMAIENQKVIETLTQEIYELAGQEFNINSPKQLGKL 539
 E PLANVLAKMEI GIKV + TL +MA +N+ +IE LTQEIY++AGQEFNINSPKQLG +
 Sbjct: 480 ELPLANVLAKMEIEGIKVNRAITQDMAEQNKVIEALTQEIYDMAGQEFNINSPKQLGSI 539

Query: 540 LFETLGLPVEMTKKTKTGYSTAVDVLERLAPISPLVTKILEYRQITKLQSTYIIGLQDYI 599
 LFE + LP+EMTKKTKTGYSTAV+VLERLAPI+P+V KIL+YRQITKLQSTY+IGLQDYI
 Sbjct: 540 LFEKMQLPLEMTKKTKTGYSTAVNVLERLAPIAPIVAKILDYRQITKLQSTYVIGLQDYI 599

Query: 600 LEDGKIHTRYVQDLTQTGRLSSSDPNLQNIPIVRLEQGRILIRKAFVPSEDNAVLLSSDYSQ 659
 L DGKIHTRYVQDLTQTGRLSS DPNLQNIPIRLEQGRILIRKAF PS ++AVLLSSDYSQ
 Sbjct: 600 LADGKIHTRYVQDLTQTGRLSSSDPNLQNIPIRLEQGRILIRKAFTPSHEDAVLLSSDYSQ 659

Query: 660 IELRVLAHISKDEHLIAAFKEGADIHTSTAMRVFGIEKPENVTPNDRRNAKAVNFGIVYG 719
 IELRVLAHIS DEHLIAAF EGADIHTSTAMRVFGI++ +VT NDRRNAKAVNFGIVYG

-1270-

Sbjct: 660 IELRVLAHISGDEHLIAAFNEGADIHTSTAMRVFGIDRAADVTANDRRNAKAVNFGIVYG 719

Query: 720 ISDFGLSHNLGIPRKLAKQYIDTYFERYPGIKNYMETVVREAKDKGYVETLFHRRRSLPD 779
ISDFGLS+NLGI RK AK YIDTYFERYPGIK YME VVREAKDKGYVETLF RRR LPD

5 Sbjct: 720 ISDFGLSNNLGITRKQAKSYIDTYFERYPGIKAYMENVVREAKDKGYVETLFKRRREL PD 779

Query: 780 INSRNFNIRQFAERTAINSPIQGSADILKIAMINLDRVLDKGGYKSKMLLQVHDEIVLE 839
INSRNFN+R FAERTAINSPIQGSADILKIAMINLD+ L GG+++KMLLQVHDEIVLE

10 Sbjct: 780 INSRNFNVRSAERTAINSPIQGSADILKIAMINLDKALQAGGFRAKMLLQVHDEIVLE 839

Query: 840 VPNEEIGAIRELVTKTMEASISLSPLIADENAGETWYEAK 880
VPN+E+ AI++LV TME+A+ L+VPL DE+ G +WYEAK

Sbjct: 840 VPNDeltaIKKLVKDTMEAAVDLAVPLCVDESTGHSWYEAK 880

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1135

A DNA sequence (GBSx1211) was identified in *S.agalactiae* <SEQ ID 3523> which encodes the amino acid sequence <SEQ ID 3524>. Analysis of this protein sequence reveals the following:

20 Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1880(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9571> which encodes amino acid sequence <SEQ ID 9572> was also identified.

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05860 GB:AP001514 unknown conserved protein [Bacillus halodurans]
Identities = 72/134 (53%), Positives = 94/134 (69%), Gaps = 3/134 (2%)

35 Query: 17 NPSDFMLKNYLTKAKTIAVVGLSDRQETAAYQVSKIMQEAGYQIIPVNPKNAGQKILGQM 76
NPSD +K L +AK IAVVGLS + +Y VS MQ AGY+IIPVNP ++LG+

Sbjct: 4 NPSDEKIQILQEAKRIAVVGLSGNPDRTSYMVSAAMQHAGYEIIPVNP--TVDEVLG EK 61

Query: 77 TYASLKDVT EHDIVNIFRSEYLPDIAREFLEVDADIFWAQLGLESQEATILKQAGHK 136
SL+D+ +DIVN+FRSE+LPD+ARE +E+ A +FWAQLGLE++EA L+Q G

40 Sbjct: 62 AVPSLQDIEGAVIDVNVFRSEHLDPDARETVEIGAPVFWAQLGLENKEAYDYLQQHGVT 121

Query: 137 QIVMNKCLKVE CQK 150
I MN+C+KVE K

45 Sbjct: 122 SI-MNRCIKVEHAK 134

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3525> which encodes the amino acid sequence <SEQ ID 3526>. Analysis of this protein sequence reveals the following:

50 Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0837(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/141 (61%), Positives = 114/141 (80%)

-1271-

Query: 11 MVYHFQNPSEDFMLKNYLTAKTIAVVGLSDRQETAAYQVSKIMQEAGYQIIPVNPKNAGQ 70
 ++Y FQNPS+ +LK YL AKTIAVVGLSDR++TAAY V+K MQ Y+IIPVNP K AGQ
 Sbjct: 1 VIYSFQNPSEDLKAYLESAKTIAVVGLSDRKDTAAYGVAKFMQAMDYRIIPVNP KLAGQ 60

Query: 71 KILGQMTYASLKDVTETHTDIVNIFRRSEYLPDIAREFLEVDADIFWAQLGLESQEAETIL 130
 ILG+ YAS+K + +DIV++FRRSE+LP++AR+FL A +FWAQLGLE+QEA+TIL
 Sbjct: 61 LILGEKVYASIKAIPEFVDIVDFRRSEFLPEVARDFLAGQAKVFWAQLGLENQEAQTIL 120

Query: 131 KQAGHKQIVMNKCLKVECQKL 151
 + AG + IVMN+CLK++ +L
 Sbjct: 121 RSAGKEAIVMNRLKIDYLOL 141

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1136

A DNA sequence (GBSx1212) was identified in *Sagalactiae* <SEQ ID 3527> which encodes the amino acid sequence <SEQ ID 3528>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3367(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9573> which encodes amino acid sequence <SEQ ID 9574> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3529> which encodes the amino acid sequence <SEQ ID 3530>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4960(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 113/151 (74%), Positives = 133/151 (87%), Gaps = 1/151 (0%)

Query: 7 MDSHSHGHRPLDAYENVLEHLREKRIRITETRKAIISYMVNSREHPSAEKIYNDLLPEYP 66
 MD HSH + LDAYENVLEHLREK IRTETRKAIISYM+ S EHPSA+KIY DL P +P
 Sbjct: 1 MDIHS+QQALDAYENVLEHLREKHIRITETRKAIISYMIQSTEHPADKIYRDLQPNFP 59

Query: 67 NMSLATVYNNLKVLDVDEGFVTELKLCNYSTTTYDFMGHQHNLNIACEDCGKIVDFVDVLDL 126
 NMSLATVYNNLKVLDVDEGFV+ELK+ N TTYDFMGHQH+N+ CE CGKI DF+DVD++
 Sbjct: 60 NMSLATVYNNLKVLDVDEGFVSELKISNDLTTYDFMGHQHVNVVCEICGKIADFMDVDVM 119

Query: 127 DISREAHQQTGFVTRVQLVAYGICPECQK 157
 DI++EAH+QTG++VTR+ ++AYGICP+CQ K
 Sbjct: 120 DIAKEAHEQTGYKVTRIPVIAYGICPDCQAK 150

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1272-

Example 1137

A DNA sequence (GBSx1213) was identified in *S.agalactiae* <SEQ ID 3531> which encodes the amino acid sequence <SEQ ID 3532>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.13 Transmembrane 16 - 32 (14 - 32)

INTEGRAL Likelihood = -1.81 Transmembrane 496 - 512 (496 - 515)

----- Final Results -----

bacterial membrane --- Certainty=0.1850(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA06650 GB:AJ005645 sdrc [Staphylococcus aureus]

Identities = 41/146 (28%), Positives = 63/146 (43%), Gaps = 13/146 (8%)

Query: 4 SQYNKWSIRRLKVGAAASVMIASGSIVALGQSHIVSAD----EMSQPKTTITAPTANTSTN 59

++ NK+SIR+ VG AS+++ + I L +A+ E++Q K TAP+ N +T

Sbjct: 16 NRLNKFISIRKYSVGTASILVGTTLIFGLSGHEAKAEHTNGELNQSKNETTAPSENKTT- 74

Query: 60 VESSTDKALKSVTTMETSEMPPK--MQNMAKVEKTSKPMVATSVRKMMATPTPVAMT- 116

D K T +++ PK M + A V++TS + T T T

Sbjct: 75 --KKVDSRQLKDNTQTATADQPKVTMSDSATVKETSSNMQSPQNATANQSTTKTSNVTTN 132

Query: 117 ---KTTSDVEVKKSTDTAFKQTVDP 139

TT +E KS T K P

Sbjct: 133 DKSSTTYSNETDKSNLTQAKDVSTTP 158

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8735> and protein <SEQ ID 8736> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7

McG: Discrim Score: -0.92

GvH: Signal Score (-7.5): -2.48

Possible site: 39

>>> Seems to have no N-terminal signal sequence

ALOM program count: 2 value: -2.13 threshold: 0.0

INTEGRAL Likelihood = -2.13 Transmembrane 16 - 32 (14 - 32)

INTEGRAL Likelihood = -1.81 Transmembrane 496 - 512 (496 - 515)

PERIPHERAL Likelihood = 7.96 402

modified ALOM score: 0.93

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.1850(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 485-489

The protein has homology with the following sequences in the databases:

D|5981|5780 leukotoxin > Insert characterized

SP|P16462|HLIA ACTAC LEUKOTOXIN. > Edit characterized

GP|141834|gb|AAA21922.1||M27399 leukotoxin (LtA) {Actinobacillus actinomycetemcomitans}

Insert characterized

Query: 210 VSLNGNTTGKEGQALLDQI|AND---KHSYQATIRVYGAKDGKVDLKNMISPKMVTINIP 266

++ NG+ + G+A +D +K + KHS + T ++ G +DL + +T P

-1273-

Sbjct: 488 ITRNGDRI-QSGKAYVDYLKKGEELAKHSDKFTKQILDPIKGNIDLSGIKGSTTLTFLNP 546

Query: 267 HITDMEVKNSLMAFKEKV-DVPAKYVSAAKAG-PFLAGVNE--TIPYEAFGGDGMLT 322
 +T E + + + E + + + K + K KG P GV + + A D +

5 Sbjct: 547 LLTAGKEERKTRQSGKYEFIGTELKVKGRTDWKVGVNSNGVYDFSNLIQHAVTRDNKVL 606

Query: 323 RLILKASEGAKWSDNGVDKNSPLL-----PLKDLTKGKYFYQVSLNGNTAGKKGQALLD 376
 L A+ GAK V S ++ + D +KG+ ++++G A K GQ ++

10 Sbjct: 607 EARLIANLGAKKDDYVFGSGSTIVNAGDGYDVVDYSKGRG-ALTIDGRNATKAGQYKVE 665

Query: 377 QIKANGSHTYQATITTIYGTGDKGV 400
 + +G+ Q T++ TK GKV

Sbjct: 666 R-DLSGTQVLQETVSKQETKRGKV 688

- 15 SEQ ID 3532 (GBS1) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 3; MW 78kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 3; MW 53kDa).

The His-fusion protein was purified as shown in Figure 189, lane 5.

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1138

A DNA sequence (GBSx1214) was identified in *S.agalactiae* <SEQ ID 3533> which encodes the amino acid sequence <SEQ ID 3534>. This protein is predicted to be response regulator (regX3). Analysis of this protein sequence reveals the following:

- 25 Possible site: 32
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
- 30 bacterial cytoplasm --- Certainty=0.3585(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 35 >GP:CAB54578 GB:AJ006397 response regulator [Streptococcus pneumoniae]
 Identities = 143/228 (62%), Positives = 183/228 (79%), Gaps = 1/228 (0%)
- Query: 1 MTQKLLLVDDDEFIIDINRRYLEQAGYEVSVAAADGIEALKEVDENRFDLIISDIMPKMD 60
 M + +LLVDE EI DI++RYL QAGY+V VA DG+EAL+ + DLII+D+MMP+MD
- 40 Sbjct: 1 MGKTILLVDDVEITDIHQRYLIQAGYQVLVAHDGLEALELFKKKPIDLIITDVMMPRMD 60
- Query: 61 GYDFISEVLVREPNOFPFLFITAKVSEPDKIYSLSMGADDFISKPFSPRELVLRVKNILRR 120
 GYD ISEV P QPFLFITAK SE DKIIY LS+GADDFI+KPFSPRELVLRV NILRR
- Sbjct: 61 GYDLISEVQYLSPEQFPFLFITAKTSEQDKIYGLSLGADDFIAKPFSPRELVLRVHNLRR 120
- 45 Query: 121 IYGNHQQSEVLTTIGDLVIDQKQRLVMVDCNTISLTNKSFDLLWILANHLNRVFSKTELYE 180
 ++ ++E+++G+L ++ V + + LT KSF+LLWILA++ RVFSKT+LYE
- Sbjct: 121 LH-RGETELISLGNLKMNHSSHEVQIGEEMDLTVKSFELLWILASNPVFSKTDLYE 179
- Query: 181 RVWGEEFLDDTINTLVNVIHALRNDLAKFSTDNTPTIKTVWGLGYKLEE 228
 ++W E+++DDTINTLVNVIHALR +LAK+S+D TPTIKTVWGLGYK+E+
- 50 Sbjct: 180 KIWKEDYDDTINTLVNVIHALRQELAKYSSDQTPTIKTVWGLGYKIEK 227

There is also homology to SEQ ID 1182.

- 55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1274-

Example 1139

A DNA sequence (GBSx1215) was identified in *S.agalactiae* <SEQ ID 3535> which encodes the amino acid sequence <SEQ ID 3536>. This protein is predicted to be histidine kinase (resE). Analysis of this protein sequence reveals the following:

```

5   Possible site: 25
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.13    Transmembrane    42 - 58 ( 33 - 65)
      INTEGRAL    Likelihood = -7.54    Transmembrane     7 - 23 (  3 - 29)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  The protein has homology with the following sequences in the GENPEPT database.
    >GP:CAB54579 GB:AJ006397 histidine kinase [Streptococcus pneumoniae]
      Identities = 190/343 (55%), Positives = 249/343 (72%)

20  Query: 1   MKLKYYIVIGYLISMLITVAGVFFGLNHMLIETRGVYYILSVTIIACIVGGIVNLFLLSS 60
      MKLK YI++GY+IS L+T+ VF+ + MLI +Y++L +TI+A +VG ++LFL
      Sbjct: 1   MKLKSYILVGYYIISTLTILVVFVAVQKMLIAKGEIYFLLGMTIVASLVGAGISLFLLLP 60

      Query: 61  VFTSLKKLKQKMKDISQRCFDTKAQICSPQEFKDLETA FNQMSSELESTFKSLNESEREK 120
      VFTSL KLK+ K ++ + F + ++ P EF+ L FN+MS +L+ +F SL ESEREK
      Sbjct: 61  VFTSLGKLKEHAKRVAADFPNLEVQGPVEFQQLGQTFNEMSHDLQVSFDSLEESEREK 120

      Query: 121 TMMLAQLSHDIKTPITSIQSTVEGILDGIISEEEVNYYLNTISRQTNRLNHLVEELSFIT 180
      +MIAQLSHDIKTPITSIQ+TVEGILDGII E E +YL TI RQT RLN LVEEL+F+T
      Sbjct: 121 GLMIAQLSHDIKTPITSIQATVEGILDGIIKESEQAHYLATIGRQTERLNKLVEELNFLT 180

30  Query: 181 LETMSDTAEPHKEETIYLDKLLIDILSEFQLVFEKENRQVMIDVAPDVSKLSSQYDKLSR 240
      L T + E +++I+LDKLLI+ +SEFQ + E+E R V + V P+ +++ Y KLSR
      Sbjct: 181 LNTARNQVETTSKDSIFLDKLLIECMSEFQFLIEQERRDVHLQVIPESARIEGDYAKLSR 240

35  Query: 241 ILLNLISNAVKYSDPGSPLTIKAYSNRQDIVIDIIDQGYGIKDEDLASIFNRLYRVESSR 300
      IL+NL+ NA KYS PG+ L + A + + I + D+G GI EDL +IF RLYRVE+SR
      Sbjct: 241 IILVNLVDNAFKYSAPGKLEVVAKLEKDQLSISVTDEGQGIAPEDLENIFKRLYRVETSR 300

40  Query: 301 NMKTGGHGLGLYIARQLAHQLNGDILVESQYQKSGKFSVLVLKL 343
      NMKTGGHGLGL IAR+LAHQL G+I V SQY GS F+LVL L
      Sbjct: 301 NMKTGGHGLGLAIARELAHQLGGEITVSSQYGLGSTFTLVNL 343

```

There is also homology to SEQ ID 1178.

A related GBS gene <SEQ ID 8737> and protein <SEQ ID 8738> were also identified. Analysis of this protein sequence reveals the following:

```

45  Lipop: Possible site: -1   Crend: 3
      McG: Discrim Score:      8.67
      GvH: Signal Score (-7.5): -5.75
      Possible site: 25
50  >>> Seems to have an uncleavable N-term signal seq
      ALOM program    count: 2 value: -9.13 threshold: 0.0
      INTEGRAL    Likelihood = -9.13    Transmembrane    42 - 58 ( 33 - 65)
      INTEGRAL    Likelihood = -7.54    Transmembrane     7 - 23 (  3 - 29)
      PERIPHERAL  Likelihood =  3.92          196
55  modified ALOM score:  2.33

      *** Reasoning Step: 3

      ----- Final Results -----
60  bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1275-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55.3/72.7% over 343aa

Streptococcus

pneumoniae

GP|5830539| histidine kinase Insert characterized

ORF00129(301 - 1332 of 1635)

GP|5830539|emb|CAB54579.1|AJ006397(1 - 344 of 350) histidine kinase {Streptococcus pneumoniae}

%Match = 34.0

%Identity = 55.2 %Similarity = 72.7

Matches = 190 Mismatches = 94 Conservative Sub.s = 60

42 72 102 132 162 192 222 252

VIWLSTKNVW*WNTAIOEP*PINHLTCFGY*OII*IVFFOKOSEFMNVSGAKNF*MTLLI*MFISMPYAMTLLNLVOTTP

282 312 342 372 402 432 462 492
QLSKQFGD*GIN*RNKMKLKYIIVIGYILISMLITVAGVFGLNHLIETRGVYYILSVTIIACIVGGIVNLFLSSVFTS
||| |:::||:| |:| : || : : || : ::| :||:| :|| : :||| |||
MKLSYILVGYYISTLLTLILVVFAVQKMLIKAEITYFLLGMTIVASLVGAGISLFLLLPVFTS
10 20 30 40 50 60

[illegible]

762 792 822 852 882 912 942 972
 ILDGIISEEEVNYLNTISRQTNRNLNHLVEELSFITLETMSDTAEPHKETIYLDKLLDILSEFQLVFEKENRQVMIDV
 ||||| | :|| | || ||||| :|:| | : ||:|:| ||||| : ||||| :|:| | | :|
 ILDGIIKESEQAHYLATIGRQTERLNKLVEELNFLTNTARNQVETTSKDSIFLDKLLIECMSEFQFLIEQERRDVHLQV
 160 170 180 190 200 210 220

1002	1032	1062	1092	1122	1152	1182	1212
APDVSKLSSQYDKLSRILLNLISNAXKYSDPGSPLTIKAYSNRQDIVIDIIDQGYGIKDEDLASIFNRLRYVESRNMKT							
: ::	:	:	:	: :	:	:	:
IPESARIEGDYAKLSRILVLNVDNAFKYSPGKTLEVVAKLEKDQLSISVTEDEGGIAPEDLENIKRLRYVETSRNMKT							
240	250	260	270	280	290	300	

1242 1272 1302 1332 1362 1392 1422 1452
GGHGLGLYIARQLAHQLNGDILVESQYQKGSFSLVLKLQK*LGII PSYFL*CFYKRLSAQ*FGKEGDRYRLIRN*RL*G
| | | | | | | | : | | | | | : | | | | | | : | | | |
GGHGLGLAIARELAHQLGGEITVSSQYGLGSTFTLVNLNLSGSENKA
320 330 340 350

SEQ ID 8738 (GBS28) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 3; MW 64kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 5; MW 38.8kDa) and in Figure 157 (lane 9-11; MW 39kDa).

GBS28-His was purified as shown in Figure 221, lane 6-7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1140

A DNA sequence (GBSx1216) was identified in *S.agalactiae* <SEQ ID 3537> which encodes the amino acid sequence <SEQ ID 3538>. Analysis of this protein sequence reveals the following:

-1276-

Possible site: 19

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -7.70 Transmembrane 125 - 141 (110 - 155)
 INTEGRAL Likelihood = -7.59 Transmembrane 38 - 54 (36 - 56)
 INTEGRAL Likelihood = -6.48 Transmembrane 146 - 162 (143 - 174)
 INTEGRAL Likelihood = -5.57 Transmembrane 72 - 88 (63 - 93)
 INTEGRAL Likelihood = -1.33 Transmembrane 229 - 245 (227 - 245)

----- Final Results -----

bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9575> which encodes amino acid sequence <SEQ ID 9576> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA79984 GB:Z21972 ORF1 [Bacillus megaterium]

Identities = 35/119 (29%), Positives = 62/119 (51%), Gaps = 15/119 (12%)

Query: 142 SSFRLLLSGNLILAPVLIVVSSLITTKAVIKLV---QQYYSYSISTLVFYTQLESGNYEG 198

+SF+L+ +++ A + + S L+ +IK + QQ++ + YT LE+

Sbjct: 105 TSFKLI-GASILQAIFIFLWSLLLIIPGIIKAIAYSQQFFL--LKDHPEYTVLEA----- 156

Query: 199 PSKVLVASRELMMNGNKLRLFLDLFSFIGWQFLTIFSFGLVYIYLLPYQTARLIFYRNI 257

+ S++ M G K + FL+ LSFIGN L +F+ G+ ++L+PY T FY +

Sbjct: 157 ----ITESKKRMKGLKWKYFLMHLFSFIGWGILCMFTLIGILLWLIPYAGTTTAAFYEEEL 211

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3539> which encodes the amino acid sequence <SEQ ID 3540>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -10.08 Transmembrane 148 - 164 (143 - 170)
 INTEGRAL Likelihood = -8.28 Transmembrane 114 - 130 (101 - 141)
 INTEGRAL Likelihood = -6.69 Transmembrane 60 - 76 (49 - 82)
 INTEGRAL Likelihood = -3.72 Transmembrane 21 - 37 (21 - 39)
 INTEGRAL Likelihood = -2.34 Transmembrane 222 - 238 (221 - 239)

----- Final Results -----

bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA79984 GB:Z21972 ORF1 [Bacillus megaterium]

Identities = 63/220 (28%), Positives = 100/220 (44%), Gaps = 31/220 (14%)

Query: 62 LGLILSLFILSASFTMI-DVVRHFRQKVSFAESTTAFSKEFFGNLLVLAITKWLFFLIWS 120

+ L+L LF+++ F +I +V+ + T + F + +A+ L S

Sbjct: 22 VSLMLLLFLINLVFPLIVEVIGSGGFSEWLMQEETPLWSDIFSMVFSIALIP---LTIS 77

Query: 121 LIWFF-----GLFIFLSGLSAFLVNAKSGSSTVISLIFLLFGAVLSLIGFGI 167

WF+ I+ G ++F + G+S + ++ L+ +L + G

Sbjct: 78 TTWFYLNLVREGNPGIPEVFFAIYKDKTSFKL---IGASILQAIFIFLWSLLLIIPG--- 131

Query: 168 YINRYAYSLSEYLLYDEVKEGTYLGAIAVIETSVAMMKGYKWLFFLQLSFTGWFLNI 227

I + AYS +LL D E T L AI S MKG KWK F + LSF GW +L +

Sbjct: 132 -IIKAIAYSQQFFLLKDH-PEYTVLEA.T--ESKKRMKGLKWKYFLMHLFSFIGWGILCM 186

Query: 228 VTFGLLNIIYLLPYFTTANVIFYDQLKKRFDKDD--PIEG 265

T G+ ++L+PY T FY++L +D DD IEG

Sbjct: 187 FTLIGILLWLIPYAGTTTAAFYEEELIVPQEDIDDDQQIEG 226

-1277-

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/254 (34%), Positives = 137/254 (53%), Gaps = 10/254 (3%)

```

Query: 16 MTNSEIKNEAKTILSNLQGNQLFLLPILLSITTLTLYISFYQYN-----NMTLLDFFVPL 70
5      M+  IK +A+  L NL GK  LFL+P LL +    I  + Y      ++L  + PL
Sbjct: 1  MSIAIKGQARDTLKNLSGKYLFLIPTLLFMFHFGIEIHQGYVLSSGIEVSLAASYFPL 60

Query: 71 PVYFFYTFLFIISVSFVMLDVVKNQKLNVRFSNTYVFSHFVWKLKLSVLVLKGLILSFFY 130
      +  +LFI+S SF M+DVV++ +  V F+++T FS  F  LL + + K L  +
10     Sbjct: 61 LLGLILSLFILSASFMTIDVVRHFRQKVSFAESTTAFSKEFFGNLLVLAITKWLFFLIWS 120

Query: 131 LLSTFGLLIISFRLLL-----SGNLILAPVLIVVSSLITTKAVIKLVQQYYSYSISTL 185
      L+  FGL I +S    L      + +++ + ++  ++++      + +YY+YS+S
15     Sbjct: 121 LIWFFGLFIFLSGLSAFLVNAKSGSSTVISLIFLLFGAVLSLIGFGIYINRYAYSLSLSEY 180

Query: 186 VFYTOLESGNYEGPSKVLVASRELMMNGNKLRLFLDLDSFIGWQFLTIFSFGLVYIYLLPY 245
      + Y +++ G Y G  V+  S  +M G K +LF L LSF GW  L I +FGL+ IYLLPY
20     Sbjct: 181 LLYDEVKEGTYLGAIAVIETSVAMMKGYKWKLFLLQSFSGWFLNIVTFGLLNLYLLPY 240

Query: 246 QTTARLIFYRNITK 259
      TTA +IFY  + K
20     Sbjct: 241 FTTANVIFYDQLKK 254

```

A related GBS gene <SEQ ID 8739> and protein <SEQ ID 8740> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 4
McG: Discrim Score: -11.32
GvH: Signal Score (~7.5): -5.39
    Possible site: 19
30  >>> Seems to have no N-terminal signal sequence
ALOM program  count: 5 value: -7.70 threshold: 0.0
    INTEGRAL    Likelihood = -7.70    Transmembrane 125 - 141 ( 110 - 155)
    INTEGRAL    Likelihood = -7.59    Transmembrane  38 -  54 (  34 -  56)
    INTEGRAL    Likelihood = -6.48    Transmembrane 146 - 162 ( 143 - 174)
35  INTEGRAL    Likelihood = -5.57    Transmembrane  72 -  88 (  63 -  93)
    INTEGRAL    Likelihood = -1.33    Transmembrane 229 - 245 ( 227 - 245)
    PERIPHERAL  Likelihood =  0.37      105
    modified ALOM score:  2.04

40  *** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

ORF00498(901 - 1071 of 1383)
EGAD|19922|20421(155 - 211 of 226)  hypothetical protein {Bacillus megaterium}
50  GP|288299|emb|CAA79984.1||Z21972 ORF1 {Bacillus megaterium} PIR|S32215|S32215 hypothetical
    protein 1 - Bacillus megaterium
    %Match = 4.8
    %Identity = 36.8  %Similarity = 61.4
    Matches = 21  Mismatches = 22  Conservative Sub.s = 14
55
    741      771      801      831      861      891      921      951
    LIIISFRLLLSGNLILAPVLIVVSSLITTKAVIKLVQQYYSYSISTLVFYTOLESGNYEGPSKVLVASRELMMNGNKLRL
                                     : : | : | | :
60  GIPEVFAIYKDGKTSFKLIGASILQAFIFLWSLLLIIPGIIKAIAYSQQFFLLKDHPEYTVLEAITESKKRMKGLKWKY
      110      120      130      140      150      160      170

    981      1011      1041      1071      1101      1131      1161      1191
    FLDDLDSFIGWQFLTIFSFGLVYIYLLPYQTTARLIFYRNITKNS*E*FLAIFVI*VLKRTYCLFDTDFRPKYPHSVDVQV
    ||: ||||| | :||: :||: || :

```

-1278-

FLMHLSTFIGWILCMFTLIGIGLLWLIPYAGTTTAAFYEELIVPQEDIDDDQQIEG
 190 200 210 220

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1141

A DNA sequence (GBSx1217) was identified in *S.agalactiae* <SEQ ID 3541> which encodes the amino acid sequence <SEQ ID 3542>. This protein is predicted to be tRNA-guanine transglycosylase (tgt). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3706(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9577> which encodes amino acid sequence <SEQ ID 9578> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14731 GB:Z99118 tRNA-guanine transglycosylase [Bacillus subtilis]
 Identities = 269/377 (71%), Positives = 320/377 (84%)

Query: 12 MTDHPIKYRLIKQEKHTGARLGEIITPHGTFPTPMFMPVGTQATVKTQSPEELKEMGSGI 71
 M + PI+Y IK+ K TGARLG++ TPHG+F TP+EMPVGT ATVKT SPEELK M +GI
 Sbjct: 1 MAEQPIRYEFIKECKQTGARLGKVHTPHGSFETPVFMPVGTLATVKTMSPEELKAMDAGI 60

Query: 72 ILSNTYHLWLRPGDELIAGGLHFKFMNWDQAILTDSGGFQVYSLADSRNITEEGVTFKN 131
 ILSNTYHLWLRPG +++ +AGGLHFKFMNWD+AILTDSGGFQV+SL+ RNI EEGV F+N
 Sbjct: 61 ILSNTYHLWLRPGQDIVKEAGGLHFKFMNWDRAILTDSGGFQVFSLSKFRNIEEGVHFRN 120

Query: 132 HINGAKMFLSPEKAISIQNNLGSDIMMSFDECPQFYQPYDYVKKSIERTSRWAERGLNAH 191
 HING K+FLSPEKA+ IQN LGSDIMM+FDECP + YDY+K+S+ERTSRWAER LNAH
 Sbjct: 121 HINGDKLFLSPEKAMEIQNALGSDIMMAFDECPPYPAEYDYMKRVERTSRWAERCLNAH 180

Query: 192 RRPHDQGLFGIVQGAGFEDLRRQSARDLVSMDFPGYSIGGLAVGETHDEMNAVLDFTVPM 251
 R +QGLFGIVQG +EDLR QSA+DL+S+DFPGY+IGGL+VGE D MN VL+FT P+
 Sbjct: 181 NRQDEQGLFGIVQGGEYEDLRTQSAKDLISLDFPGYAIGGLSVGEPKDVMNRVLEFTTPL 240

Query: 252 LPNDKPRYLMGVGAPDSLIDAVIRGVDMFDCVLPTRIARNGTCMTSQGRLVVKNKAFED 311
 LP DKPRYLMGVG+PD+LID IRGVDMFDCVLPTRIARNGT T++GRL +KNKF D
 Sbjct: 241 LPKDKPRYLMGVGSPDALIDGAIRGVDMFDCVLPTRIARNGTVFTAEGRLNMKNKAFERD 300

Query: 312 FTPLDPNCDCYTCKNYTRAYIRHLKADETFGIRLTSYHNLYFLVNLMKDVRQAIMDDNL 371
 F P+D CDCYTCKNYTRAYIRHL++ +ETFG+RLT+YHNL+FL++LM+ VRQAI +D L
 Sbjct: 301 FRPIDEECDYTCKNYTRAYIRHLIRCNETFGRLRTTYHNLHFLHLMEQVRQAIREDRL 360

Query: 372 LEFRQDFMERYGYGMNN 388
 +FR++F ERYGY N
 Sbjct: 361 GDFREEFFERYGYNKPN 377

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3543> which encodes the amino acid sequence <SEQ ID 3544>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>

-1279-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 351/380 (92%), Positives = 368/380 (96%)

Query: 12 MTDHPKPYRLIKQEKHTGARLGEIITPHGTFPTPMFMPVGTQATVKTSPEELKEMGSGI 71
 MTD+PIKYRLIK EKHTGARLGEIITPHGTFPTPMFMPVGTQATVKTSPEELK +GSGI
 10 Sbjct: 1 MTDYPKPYRLIKAEKHTGARLGEIITPHGTFPTPMFMPVGTQATVKTSPEELKAIGSGI 60

Query: 72 ILSNTYHLWLRPGDELIAGGLHKKFMNWDQAILTDSGGFQVYSLADSRNITEEGVTFKN 131
 ILSNTYHLWLRPGDELIA++GGLHKKFMNWDQ ILTDSGGFQVYSLADSRNITEEGVTFKN
 15 Sbjct: 61 ILSNTYHLWLRPGDELIARSGGLHKKFMNWDQPILTDSGGFQVYSLADSRNITEEGVTFKN 120

Query: 132 HLNGAKMFLSPEKAISIQNNLGSDIMMSFDECPQFYQPYDYVKKSIERTSRWAERGLNAH 191
 HLNG+KMFLSPEKAISIQNNLGSDIMMSFDECPQFYQPYDYVKKSIERTSRWAERGL AH
 20 Sbjct: 121 HLNGSKMFLSPEKAISIQNNLGSDIMMSFDECPQFYQPYDYVKKSIERTSRWAERGLKAH 180

Query: 192 RRPHDQGLFGIVQGAGFEDLRRQSARDLVSMDFPGYSIGGLAVGETHDEMNAVLDFTVPM 251
 RRPHDQGLFGIVQGAGFEDLRRQSA DLV+MDFPGYSIGGLAVGE+H+EMNAVLDFT P+
 25 Sbjct: 181 RRPHDQGLFGIVQGAGFEDLRRQSAADLVAMDFPGYSIGGLAVGESHEEMNAVLDFTTPL 240

Query: 252 LPNDKPRYLMGVGAPDSLIDAVIRGVDMFDCVLPTRIARNGTCMTSQGRLVVKNAKFAED 311
 LP +KPRYLMGVGAPDSLID VIRGVDMFDCVLPTRIARNGTCMTS+GRLV+KNKFAED
 30 Sbjct: 241 LPENKPRYLMGVGAPDSLIDGVIRGVDMFDCVLPTRIARNGTCMTSEGRLVIKNKFAED 300

Query: 312 FTPLDPNCDCYTCCKNYTRAYIRHLLKADETFGIRLTSYHNLYFLVNLMKDVRQAIMDDNL 371
 FTPLD +CDCYTC+NY+RAYIRHLLKADETFGIRLTSYHNLYFLVNLMK VRQAIMDDNL
 35 Sbjct: 301 FTPLDHDCCYTCQNYRAYIRHLLKADETFGIRLTSYHNLYFLVNLMKKVRQAIMDDNL 360

Query: 372 LEFRQDFMERYGYGMNNRNF 391
 LEFRQDF+ERYGY +NRNF
 40 Sbjct: 361 LEFRQDFLERYGYNKSNNRNF 380

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1142

A DNA sequence (GBSx1218) was identified in *S. galactiae* <SEQ ID 3545> which encodes the amino acid sequence <SEQ ID 3546>. Analysis of this protein sequence reveals the following:

40 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2479 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9303> which encodes amino acid sequence <SEQ ID 9304> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10795> which encodes amino acid sequence <SEQ ID 10796> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16256 GB:Z99164 hypothetical protein [Schizosaccharomyces pombel]
 Identities = 42/91 (46%), Positives = 62/91 (67%), Gaps = 3/91 (3%)

55 Query: 6 FGIGLDSSSRCYHYHTKLDIVALKCAVCQKYYACYKCHDALEEHCFAA-TKSDETFP-VL 63
 +G +D+ +RC+HYH+K D+VAL+C C+K+YAC++CHD L H F K+ P V+
 Sbjct: 13 YGKLVNDNETRCFHYHSHKADVVVALRCGQCEKFYACFQCHDELNTHPFLPWRKAKFHIPCVI 72

-1280-

Query: 64 CGSCRQMLTLKEYK-TGFCPYCRMLFNPNCQ 93
 CG+C+ LT++EY+ T C YC FNP C+
 Sbjet: 73 CGACKNSLTVVEYRSTVHCKYCNHFPNPKCK 103

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3547> which encodes the amino acid sequence <SEQ ID 3548>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2769(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 55/93 (59%), Positives = 62/93 (66%)

Query: 2 MQEYFGIGLDSSSRCYHYHTKLDIVALKCAVCQKYACVKCHDALEEHCFATKSDETFP 61
 M + FGI LD RC HYHT LDIV LKCA CQ YYACY CHD L +H F T ET P
 Sbjet: 1 MTD CFGIDL DQEYRCLHYHTPLDIVGLKASCQTYACVHCHDQLTDHAFVPTGHQETSP 60

Query: 62 VLCGSCRQMLTLKEYKTGFCPYCRMLFNPNCQR 94
 V+CG CR++L+ EY G CPYC+ FNP C R
 Sbjet: 61 VICGHCRLLSRAEYGC GCCPYCQSPFNPACHR 93

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1143

A DNA sequence (GBSx1219) was identified in *S.agalactiae* <SEQ ID 3549> which encodes the amino acid sequence <SEQ ID 3550>. This protein is predicted to be transport protein. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.45 Transmembrane 300 - 316 (292 - 321)
 INTEGRAL Likelihood = -1.17 Transmembrane 265 - 281 (265 - 281)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10113> which encodes amino acid sequence <SEQ ID 10114> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF12002 GB:AE002075 transport protein, putative [Deinococcus radiodurans]
 Identities = 108/295 (36%), Positives = 174/295 (58%), Gaps = 4/295 (1%)

Query: 31 GAWINLVNPSQEESEQVADQFGIDIDDLRAPLDVEETSRISEDDYTLVIVDVPTYEERN 90
 G WI+ P+ EE +V+ + G+++D L+ PLD +E SR ED L+I+ +
 Sbjet: 21 GCWIDAAAPTTEELARVSRETGLELDYLYKPLDPDERSRFRERDQQLLIIMQTSYRLAED 80

Query: 91 NKSYYMTIPMGIIIVTDNAVITTC-LEHLTLFDHFYRRRVKNFYTFMKTRFVFQLLYRNAE 149
 + Y T+P+GI+ TD+ ++T C LE + V+ T K R QL RNA+
 Sbjet: 81 SDIPYDTPVLGILHTDHCILVTVCLEENPVVKDVSGLVRRVSTVKKNRLTLQLFLRNAQ 140

-1281-

Query: 150 LYLQALRTIDRQSDKIEAQLESATRNEQLIDMMELEKSIVYLKASLKFNERIVKKLTSST 209
 +L +R I+++ D IE ++E+ATRN +L+D+++LEKS+VY LK NE +++++
 Sbjct: 141 RFLIDVRQINKRVDIEDKMNATRNRELLDLLKLEKSLVYFITGLKANEAMMERVKRDR 200

5 Query: 210 SSLKKYIEDEDLLEDTLIETQQAIEMANIYENVLNAMTETTASIIGNNQNTIMKTLALVT 269
 + Y ED +LL+D LIE QAIEMA+I N+L +M AS+I NN N ++K L + T
 Sbjct: 201 I-FEMYEEDSELDDVLIENLQAIEMASIASNILTSMAGAFASVINNNVNQVVKVLTVT 259

10 Query: 270 MTLDIPTVIFSAYGMNFQNNWMPNLGLAHGFIYVLLAFLMSSFVVFYFIRKKWF 324
 + + IPT++ +GMN + +P + +GF V+ +A ++S + F F R K F
 Sbjct: 260 ILVAIPTLVSGFFGMNVEG--LPFSDSPYGFVLVMTVAMGIASLLAFLFYRWKVF 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 715> which encodes the amino acid sequence <SEQ ID 716>. Analysis of this protein sequence reveals the following:

15 Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.81 Transmembrane 293 - 309 (288 - 311)
 INTEGRAL Likelihood = -1.28 Transmembrane 255 - 271 (255 - 271)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 272/314 (86%), Positives = 296/314 (93%)

Query: 11 MKQMFLSTAIEFKEIETFEPGAWINLVNPSQESEQVADQFGIDIDDLRAPLDVEETSRI 70
 MKQMFLS+AIEFKEIETFEPGAWI LVNPSQEEs ++ADQF IDI DLRAPLDVEETSRI
 30 Sbjct: 1 MKQMFLSSAIEFKEIETFEPGAWIKLNVNPSQEEsMKIADQFNIDISDLRAPLDVEETSRI 60

Query: 71 SVEDDYTLVIVDVPTYEERNNKSYYMTIPMGIIIVTDNAVITTCLEHLTLFDHFYRRRVKN 130
 +VEDDYTL+IVDVP YEERNNKSYY+T+P+GIIVT+NAVITTCLEHLTLFDHF+ RRVKN
 35 Sbjct: 61 AVEDDYTLIIVDVPIYEERNNKSYYITMPLGIIVTENAVITTCLEHDMTLFDHFHNRVKN 120

Query: 131 FYTFMKTRFVFQQLLYRNAELYQALRTIDRQSDKIEAQLESATRNEQLIDMMELEKSIVY 190
 FYTFMKTRFVFQ+LYRNAEL+L ALRTIDRQS+++EAQLE+ATRNE+LIDMMELEKSIVY
 Sbjct: 121 FYTFMKTRFVFQILYRNAELFLTALRTIDRQSERLEAQLEAATRNEELIDMMELEKSIVY 180

40 Query: 191 LKASLKFNERIVKKLTSSTSSLKKYIEDEDLLEDTLIETQQAIEMANIYENVLNAMTETT 250
 LKASLKFNERIVKKL+SSTSSLKKYIEDEDLLEDTLIETQQAIEMA IYENVLNAMTETT
 Sbjct: 181 LKASLKFNERIVKKLSSSTSSLKKYIEDEDLLEDTLIETQQAIEMAGIYENVLNAMTETT 240

45 Query: 251 ASIIGNNQNTIMKTLALVTMTLDIPTVIFSAYGMNFQNNWMPNLGLAHGFIYVLLAFLM 310
 ASII NNQNTIMKTLAL+TM LDIPTVIFSAYGMNFQNNW+PLNGL H F Y+ L+A L+
 Sbjct: 241 ASIINNQNNTIMKTLALMTALDIPTVIFSAYGMNFQNNWPLNGLHAFWYITLIAMLL 300

Query: 311 SSFVVFYFIRKKWF 324
 SSFVV YFIRKKWF
 50 Sbjct: 301 SSFVVIYFIRKKWF 314

SEQ ID 3550 (GBS257) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 3; MW 35kDa), in Figure 169 (lane 9 & 10; MW 50kDa) and in Figure 239 (lane 2; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 6; MW 60kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1282-

Example 1144

A DNA sequence (GBSx1220) was identified in *S.agalactiae* <SEQ ID 3551> which encodes the amino acid sequence <SEQ ID 3552>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.26	Transmembrane	158 - 174 (151 - 182)
INTEGRAL	Likelihood = -6.37	Transmembrane	93 - 109 (91 - 111)
INTEGRAL	Likelihood = -5.68	Transmembrane	188 - 204 (184 - 205)
INTEGRAL	Likelihood = -0.85	Transmembrane	118 - 134 (118 - 134)

----- Final Results -----

bacterial membrane	---	Certainty=0.5904(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3553> which encodes the amino acid sequence <SEQ ID 3554>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.95	Transmembrane	92 - 108 (88 - 110)
INTEGRAL	Likelihood = -6.69	Transmembrane	153 - 169 (151 - 177)
INTEGRAL	Likelihood = -2.34	Transmembrane	183 - 199 (183 - 200)

----- Final Results -----

bacterial membrane	---	Certainty=0.3781(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 135/217 (62%), Positives = 167/217 (76%), Gaps = 1/217 (0%)

```

Query: 1  MTLQDLTKKNQEFVHIATNQLLADGKSDAEIKAILLEEHLPEIIDNQKKGITARSLLGAPT 60
          M LQ+LTKKNQEF+H ATN+L+ DGKSD +IK ILEE +P I++NQKKG+TAR+LLG PT
Sbjct: 1  MELQELTKKNQEFIHTATNKLIQDGKSDEDIKLILEEAIPAILENQKKGVTARNLLGTPT 60

Query: 61 TWAASFTERPEDKARVSVQKNTNPWLMWLDTSLLFLGLVTALNGLMLLFGQSNVNTGLIS 120
          WAASF++ P KA      KNTNPWLMWLDTSLLF+G+V LNG+M F +  TGLIS
Sbjct: 61 AWAASFSDPSQKA-AETDKNTNPWLMWLDTSLLFIGIVALNGIMTFENTNATVTGLIS 119

Query: 121 ILTLGFGGGAAMYVTTYTYIRHMGKPKSERPGWLKSFVAVLALVMLVWFALFAVVPLLPAT 180
          +L LGFGGGA+MY TYY+IYRH+GK KS RP W K A L+L ML+W AL++  LP +
Sbjct: 120 LLALGFGGASMYATYYFIYRHLGKDKSLRPSWFKIIAALSLAMLIWIALYSATAFLPTS 179

Query: 181 INPKLPEVVLFIIALASFGLRFYLQKRYNIQSSMAPV 217
          +NP+LP + L II  S LR+YLQKRYNIQ++M+PV
Sbjct: 180 LNPQLPPLALLIIGVSLALRYYLQKRYNIQNTMSPV 216

```

A related GBS gene <SEQ ID 10787> and protein <SEQ ID 10788> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: -9.94

GvH: Signal Score (-7.5): -3.66

Possible site: 29

>>> Seems to have no N-terminal signal sequence

ALOM program count: 4 value: -12.26 threshold: 0.0

INTEGRAL	Likelihood = -12.26	Transmembrane	158 - 174 (151 - 182)
----------	---------------------	---------------	------------------------

-1283-

INTEGRAL Likelihood = -6.37 Transmembrane 93 - 109 (91 - 111)
 INTEGRAL Likelihood = -5.68 Transmembrane 188 - 204 (184 - 205)
 INTEGRAL Likelihood = -0.85 Transmembrane 118 - 134 (118 - 134)
 PERIPHERAL Likelihood = 8.43 50
 modified ALOM score: 2.95

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1145

A DNA sequence (GBSx1221) was identified in *S.agalactiae* <SEQ ID 3555> which encodes the amino acid sequence <SEQ ID 3556>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1348(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1146

A DNA sequence (GBSx1222) was identified in *S.agalactiae* <SEQ ID 3557> which encodes the amino acid sequence <SEQ ID 3558>. This protein is predicted to be excinuclease ABC (uvrA). Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10111> which encodes amino acid sequence <SEQ ID 10112> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC67271 GB:AF017113 excinuclease ABC subunit A [Bacillus subtilis]
 Identities = 642/940 (68%), Positives = 785/940 (83%), Gaps = 3/940 (0%)

Query: 9 DKLMIKGAHNLKNIISVDIPRDKLVVVTGLSGSGKSSLAFTTIYAEQORRYVESLSAYA 68
 D++ ++GAHNLKNI V IPRD+LVVVTGLSGSGKSSLAFTTIYAEQORRYVESLSAYA
 Sbjct: 4 DRIEVKGAHNLKNIDVTIPRDLVVTGLSGSGKSSLAFTTIYAEQORRYVESLSAYA 63

-1284-

Query: 69 RQFLGNMEKPDVDSIDGLSPAISIDQKTTSTKNPRSTVGTVTINDYLRLLYARVGTPYCI 128
 RQFLG M+KPDVD+I+GLSPAISIDQKTTST+KNPRSTVGTVTETI DYLRLLYARVG P+C
 Sbjct: 64 RQFLGQMDKPDVDAIEGLSPAISIDQKTTSTSRNPRSTVGTVTETIYDYLRLLYARVGKPHCP 123

5 Query: 129 NGHGAITASSVEQIVDKVLALPERTKMQILAPIIRRKKGQHKSTFEKIQKDGYYVRVRIDG 188
 IT+ ++EQ+VD++L PERTK+Q+LAPI+ +KG H E+I+K GYVRVRIDG
 Sbjct: 124 EHGIEITSQTIEQMVDRILEYPERTKIQVLAPIVSGRKGAVKVKLEQIRKQGYVRVRIDG 183

10 Query: 189 DIHDVTEVPELSKSKMHNIDIVDRLINKEGIRSRLFDSVEAALRLSDGYVVIDTMDGNE 248
 ++ ++++ EL K+K H+I++V+DR++ KEG+ +RL DS+E ALRL +G V+ID + E
 Sbjct: 184 EMAELSDDIELEKNKKSIEVVIDRIVVKEGVAARLSDSLETALRLGEGRVMIDVIGEEE 243

15 Query: 249 LLFSEHYSCPECFTVPELEPRIFSFNAPFGSCPTCDGLGIKLEVDIDLVIPDRSKTLRE 308
 L+PSEH++CP CGF++ ELEPRIFSFN+PFG+CPTCDGLG+KLEVD DLVIP++ +L+E
 Sbjct: 244 LMFSEHHACPHCGFSIGELEPRIFSFPFGACPTCDGLGMKLEVDADLVI PNQDL SLKE 303

20 Query: 309 GALVPWNPISSNYPTMLEQAMTQFGVMDTPFEKLSKAEQDLALYSGSGEREFHFHYIND 368
 A+ PW PISS YYP +LE T +G+DMD P + L K + D LYSGS+ +F Y ND
 Sbjct: 304 NAVAPWTPISSQYYPQLLEAVCTHYGIDMDVPVKDLPKHQLDKVLVYSGGDDLIYFRYEND 363

25 Query: 369 FGGERNIDLFPFEGVVNNINRRYHETNSDYTRNVNREYMNELKCNCTCHGYRLNDQALCVRV 428
 FG R ++ FEGV+ NI RRY ET SD+ R M +YM++ C TC GYRL +AL V +
 Sbjct: 364 FGQIREGEIQFEGVLRNIERRYKETGSDFIREQMEQYMSQKSCPTCKGYRLKKEALAVLI 423

30 Query: 429 GGEGLNIGQVSDLSIADHLELLETLRSSNEQLIARPIIKEIHDRLSFLNNVGLNYLNL 488
 +G +IG++++LS+AD L + L LS + IA I++EI +RLSFL+ VGL+YL L
 Sbjct: 424 ---DGRHIGKITELSVADALAFFKDLTLSEKDMQIANLILREIVERLSFLDKVGLDYLTL 480

35 Query: 489 SRSAGTSLGGESQRIRLATQIGSNLSGVLYVLDEPSIGLHQDNDRLIDSLKKMRDLGNT 548
 SR+AGTSLSGGE+QRIRLATQIGS LSGVLY+LDEPSIGLHQDNDRLI +LK MRDLGNT
 Sbjct: 481 SRAAGTSLSGGEAQRIRLATQIGSRLSGVLYILDEPSIGLHQDNDRLISALKNMRLGNT 540

40 Query: 549 LIVVEHDEDTMMAADWLIDVGPAGAGFGEIVASGTPKQVAKNTKSITGQYLSGKKVIPV 608
 LIVVEHDEDTMMAAD+LID+GPGAG GG+++++GTP++V ++ S+TG YLSGKK IP+
 Sbjct: 541 LIVVEHDEDTMMAADYLIDIGPGAGIHGGQVISAGTPEEVMEPDNSLTGSYLSGKKFIPL 600

45 Query: 609 PSERRVGNRFLFIKGAENNLQNLVDKFPPLGKFIATVGVSGSGKSTLINSILKKAQAQK 668
 P ERR +GR++EIKGA+ENNL+ ++ KPPLG F AVTGVSGSGKSTL+N IL KA+AQK
 Sbjct: 601 PPERRKPDGRYIEIKGASENNLKKVNAKFPLGTFTAVTGVSGSGKSTLVNEILHKALAQK 660

50 Query: 669 LNRNSDKPGKYVSLIEGIEYVDRLIDIDQSPIGRTPRSPATYTGVFDDIRDLEAQTNEAK 728
 L++ KPG + ++G++++D++IDIDQ+PIGRTPRSPATYTGVFDDIRD+FAQTNEAK
 Sbjct: 661 LHKAKAKPGSHKEIKGLDHLKVIDIDQAPIGRTPRSPATYTGVFDDIRDVFAQTNEAK 720

55 Query: 729 IRGYKKGRFSFNVKGGRCESCGDGIIKIEMHFLPDVYVPCVCHGTRYNSETLEVHYKE 788
 +RGYKKGRFSFNVKGGRCE+C GDGIIKIEMHFLPDVYVPCVCHG RYN ETLEV YK
 Sbjct: 721 VRGYKKGRFSFNVKGGRCEACRGGDIIKIEMHFLPDVYVPCVCHGKRYNRETLEVITYKG 780

60 Query: 789 KNIAQILDMTVNDAVTFFAAIPKIARKLQTIKDVGLGYVTLGQPATTLSGGEAQRMKLAS 848
 K+I+ +LDMTV DA++FF IPKI RKLQT+ DVGLGY+TLGQPATTLSGGEAQR+KLAS
 Sbjct: 781 KSISDVLDMTVEDALSFFENIPKIKRKLQTLVDVGLGYITLGQPATTLSGGEAQRVKLAS 840

Query: 849 ELHKRSTGKSLYLDEPTTGLHADDIARLLKVLDRFVDDGNTVLVIEHNLDVIKTADHII 908
 ELHKRSTG++LYILDEPTTGLH DDIA RL L VL R VD+G+TVLVIEHNLD+IKTAD+I+
 Sbjct: 841 ELHKRSTGRTLYILDEPTTGLHVDDIARLLVVLQRLVDNGDTVLVIEHNLDIIKTADYIV 900

Query: 909 DLGPEGGIGGGQIVAIGTPEEVAENPKSYTGYYLKEKLAR 948
 DLGPEGG GGG IVA GTPEE+ E +SYTG YLK + R
 Sbjct: 901 DLGPEGGAGGTIVASGTPEEITEVEESYTGRYLKPVIER 940

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3559> which encodes the amino acid sequence <SEQ ID 3560>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1285-

bacterial cytoplasm --- Certainty=0.1138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 835/940 (88%), Positives = 896/940 (94%)

Query: 7 MQDKLMIRGARAHNLKNI SVDIPRDKLVVVVTGLSGSGKSSLAFTDIYAEGQRRYVESLSA 66
 MQ+K++I GARAHNLKNI V+IPRDKLVVVVTGLSGSGKSSLAFTDIYAEGQRRYVESLSA
 10 Sbjct: 11 MQNKIIHGHARAHNLKNIDVEIPRDKLVVVVTGLSGSGKSSLAFTDIYAEGQRRYVESLSA 70

Query: 67 YARQFLGNMEKPDVDSIDGLSPAISIDQKTTSKNPRSTVGTVTEINDYLRLLYARVGTPY 126
 YARQFLGNMEKPDVDSIDGLSPAISIDQKTTSKNPRSTVGTVTEINDYLRLLYARVGTPY
 15 Sbjct: 71 YARQFLGNMEKPDVDSIDGLSPAISIDQKTTSKNPRSTVGTVTEINDYLRLLYARVGTPY 130

Query: 127 CINGHGAITASSVEQIVDKVLALPERTKMQILAPIIRKKKGQHKSTFEKIQKDG YVRVRI 186
 CINGHGAITASS EQIV++VLALPERT+MQILAP++RRKKGQHK+ FEKIQKDG YVRVRI
 20 Sbjct: 131 CINGHGAITASSAEQIVEQVLALPERTRMQILAPVVRKKKGQHKTVFEKIQKDG YVRVRI 190

Query: 187 DGDIDHVTVEPELSKSKMHNI DIVDRLINKEGIRSLFDSVEAALRLSDGYVVIDTMDG 246
 DGDIDHVTVEPELSKSKMHNI++V+DRL+NK+GIRSLFDSVEAALRL DGY++IDTMDG
 25 Sbjct: 191 DGDIDHVTVEPELSKSKMHNI EVVIDRLVNKDGIRSLFDSVEAALRLSDGYLMIDTMDG 250

Query: 247 NELLFSEHYSCPCGFTVPELEPRLFSFNAPFGSCPTCDGLGIKLEVDIDLVIPDRSKTL 306
 NELLFSEHYSCP CGFTVPELEPRLFSFNAPFGSCPTCDGLGIKLEVD+DLV+PD SK+L
 30 Sbjct: 251 NELLFSEHYSCPVCGFTVPELEPRLFSFNAPFGSCPTCDGLGIKLEVDLDLVDPDSKSL 310

Query: 307 REGALVPWNPISSNYPTMLEQAMTQFGVDM DTPFEKLSKAEQDLALYSGGEREFH FHYI 366
 REGAL VWNPISSNYPTMLEQAM FGVDMDTPFE L++ E+DL LYSGG+REFH FHYI
 35 Sbjct: 311 REGALAPWNPISSNYPTMLEQAMASFGVDM DTPFEALTEERDLVLYSGGDREFH FHYI 370

Query: 367 NDFGGERNIDLPFEGVVMNINRRYHETNSDYTRNVMREYMNELKCN TCHGYRLNDQALCV 426
 NDFGGERNID+PFEGVV N+NRRYHETNSDYTRNVMR YMNEL C TCHGYRLNDQALCV
 40 Sbjct: 371 NDFGGERNIDLPFEGVVMNINRRYHETNSDYTRNVMRGYMNELTCATCHGYRLNDQALCV 430

Query: 427 RVGGEEGLNIGQVSDLSIADHLELLETIRLSSNEQLIARPIIKEIHDRLSFLNNVGLNYL 486
 VGGEEG +IGQ+S+LSIADHL+LLE L L+ NE IA+PI+KEIHDRL+FLNNVGLNYL
 45 Sbjct: 431 HVGGEETHIGQISELSIADHLQLLELELTENESTIAKPIVKEIHDRLTFLNNVGLNYL 490

Query: 487 NLSRSAGTSLGGESQRI RLATQIGSNL SGVLYVLDEPSIGLHQ RDNDRLIDS LKKMRDLG 546
 LSR+AGTSLGGESQRI RLATQIGSNL SGVLY+LDEPSIGLHQ RDNDRLI+SLKKMRDLG
 50 Sbjct: 491 TLSRAAGTSLGGESQRI RLATQIGSNL SGVLYLDEPSIGLHQ RDNDRLIES LKKMRDLG 550

Query: 547 NTLIVVEHDEDTMMADWLIDVGP GAGAFGGEIVASGTPKQVAKNTKSITGQYLSGKKVI 606
 NTLIVVEHDEDTMM ADWLIDVGP GAG FGGEI ASGTPKQVAKN KSITGQYLSGKK I
 55 Sbjct: 551 NTLIVVEHDEDTMMQADWLIDVGP GAGEFGGEITASGTPKQVAKNKSITGQYLSGKKFI 610

Query: 607 PVPSEERRVGNRFL EIKGAAENNLQNL DVKFLGKFIAVTGVSGSGKSTLINSILKKAVA 666
 PVP ERR GNGRF+EIKGAA+NNLQ+LDV+FPLGKFIAVTGVSGSGKSTL+NSILKKAVA
 60 Sbjct: 611 PVPLERRSGNGRFIEIKGAAQNNLQSLDVRFP LKGFIAVTGVSGSGKSTLVNSILKKAVA 670

Query: 667 QKLNRSNDKPGKYVSLEGIEYVDRLIDIDQSPIGRTPRSNPATYTG VFDDIRDLFAQTNE 726
 QKLNRN+DKPGKY S+ GIE+++RLIDIDQSPIGRTPRSNPATYTG VFDDIRDLFAQTNE
 65 Sbjct: 671 QKLNRNADKPGKYHSISGIEHIERLIDIDQSPIGRTPRSNPATYTG VFDDIRDLFAQTNE 730

Query: 727 AKIRGYKKGRFSFNVKGRCESCSGDGIIKIE MHFLPDVYVPCVCHGTRYNSETLEVHY 786
 AKIRGYKKGRFSFNVKGRCESCSGDGIIKIE MHFLPDVYVPCVCHG RYNSETLEVHY
 70 Sbjct: 731 AKIRGYKKGRFSFNVKGRCESCSGDGIIKIE MHFLPDVYVPCVCHGRRYNSETLEVHY 790

Query: 787 KEKNIAQILDMTVNDAVTFFAAIPK IARKLQTIKDVGLGYVTLGQPATTL SGGEAQRMKL 846
 K KNIA++LDMTV+DA+ FF+AIPK IARK+QTIKDVGLGYVTLGQPATTL SGGEAQRMKL
 75 Sbjct: 791 KGKNIAEVLDMTVDDALVFFSAIPK IARKIQTIKDVGLGYVTLGQPATTL SGGEAQRMKL 850

Query: 847 ASELHKRSTGKSLYILDEPTTGLHADDIARLLKVLDRFVDDGNTVLVIEHNL DVIKTADH 906
 ASELHKRSTGKSLYILDEPTTGLH DDIA RLLKVL+RFVDDGNTVLVIEHNL DVIK+ADH
 80 Sbjct: 851 ASELHKRSTGKSLYILDEPTTGLHTDDIARLLKVLDRFVDDGNTVLVIEHNL DVIKSADH 910

-1286-

Query: 907 IIDLGPEGGIGGGQIVAIGTPPEEVAENPKSYTGYLKEKL 946
 IIDLGPEGG GGGQIVA GTPEEVA+ +SYTG+YLK KL
 Sbjct: 911 IIDLGPEGGDGGGQIVATGTPEEVAQVKESYTGHYLKVKL 950

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1147

A DNA sequence (GBSx1223) was identified in *S.agalactiae* <SEQ ID 3561> which encodes the amino acid sequence <SEQ ID 3562>. Analysis of this protein sequence reveals the following:

10 Possible site: 60
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.40	Transmembrane	471 - 487 (463 - 490)
INTEGRAL	Likelihood = -9.29	Transmembrane	246 - 262 (242 - 264)
INTEGRAL	Likelihood = -7.27	Transmembrane	183 - 199 (178 - 207)
15 INTEGRAL	Likelihood = -5.41	Transmembrane	351 - 367 (349 - 370)
INTEGRAL	Likelihood = -4.41	Transmembrane	87 - 103 (83 - 107)
INTEGRAL	Likelihood = -3.24	Transmembrane	375 - 391 (374 - 392)
INTEGRAL	Likelihood = -2.97	Transmembrane	17 - 33 (16 - 35)
INTEGRAL	Likelihood = -2.28	Transmembrane	420 - 436 (420 - 438)
20 INTEGRAL	Likelihood = -1.97	Transmembrane	320 - 336 (320 - 337)
INTEGRAL	Likelihood = -1.75	Transmembrane	214 - 230 (214 - 230)
INTEGRAL	Likelihood = -1.75	Transmembrane	288 - 304 (288 - 304)
INTEGRAL	Likelihood = -1.70	Transmembrane	110 - 126 (110 - 126)
25 INTEGRAL	Likelihood = -0.69	Transmembrane	152 - 168 (151 - 168)

----- Final Results -----

bacterial membrane	---	Certainty=0.5161(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12192 GB:Z99106 similar to multidrug resistance protein [Bacillus subtilis]
 Identities = 198/481 (41%), Positives = 300/481 (62%), Gaps = 24/481 (4%)

35 Query: 9 IHGKPYNRTAMITLLLIATFAGVLNQTS LGTAIPTLMNSFNISLSTAQQATTWFLLANGI 68
 I KP+NR+ ++ +LL F +LNQT L TA+P +M FN+ + AQ TT F+L NGI
 Sbjct: 5 IEQKPFNRSVIVGILLAGAFVAILNQTLITLITLPHIMRDFNVNDANQAQWLTTSFMLTNGI 64

40 Query: 69 MIPVSAYLATRFSTKWLIVTSYVLLIGLLMTTLAPTSNWNFLVGRIIQAISVGISMPL 128
 +IP++A+L +F+++ L +T+ + G ++ AP N+ + L RIIQA GI MPL
 Sbjct: 65 LIPITAFLEKFTSRALLITAMSIFTAGTVVGAFAP--NFPVLLTARI IQAAGAGIMMPL 122

45 Query: 129 MQVVMVNVFPPEORGAAMGLNLVVG LAPAIGPTLAGWILKQEFHFAGHDLTWRAIFLLP 188
 MQ V + +FP E+RG AMG+ GLV+ APAIGPTL+GW ++ +WR++F +
 Sbjct: 123 MQTVFLTIFPIEKRGQAMGMVLVISFAPAIGPTLSGWAVEA-----FSWRS LFYII 174

Query: 189 LLILTVTTILSPFVLKDVVDNKS VKLEVPSLILSIIGFGSFLWGFTNVATYGWGDIGYVI 248

-1287-

Query: 429 LSSVAQNIITNNKPSKDLLTMNPLKYANQMLNASLDGFHVSFAIGFVFAVLGLLVSLFLRK 489
 L SV N + + +A+L G + +F + V A++G L+S I+K
 Sbjct: 414 LVSVMSNQAAH-----AGTTNVKHAALHGMNAAFIVAVALVGFLLSFTLKK 461

5 There is also homology to SEQ ID 46.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1148

10 A DNA sequence (GBSx1224) was identified in *S.agalactiae* <SEQ ID 3563> which encodes the amino acid sequence <SEQ ID 3564>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -8.81 Transmembrane 8 - 24 (5 - 30)
 INTEGRAL Likelihood = -7.32 Transmembrane 36 - 52 (31 - 54)
 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 10109> which encodes amino acid sequence <SEQ ID 10110> was also identified.

A related GBS gene <SEQ ID 8743> and protein <SEQ ID 8744> were also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 9.52
 GvH: Signal Score (-7.5): -3.4
 Possible site: 22
 >>> Seems to have an uncleavable N-term signal seq
 30 ALOM program count: 1 value: -7.32 threshold: 0.0
 INTEGRAL Likelihood = -7.32 Transmembrane 11 - 27 (6 - 29)
 PERIPHERAL Likelihood = 11.19 130
 modified ALOM score: 1.96
 35 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 SEQ ID 8744 (GBS29) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 2; MW 25.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 6; MW 51kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1149

A DNA sequence (GBSx1225) was identified in *S.agalactiae* <SEQ ID 3565> which encodes the amino acid sequence <SEQ ID 3566>. This protein is predicted to be aminopeptidase P (pepQ). Analysis of this protein sequence reveals the following:

```

5   Possible site: 41
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.0724(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAA70068 GB:Y08842 aminopeptidase P [Lactococcus lactis]
      Identities = 44/126 (34%), Positives = 78/126 (60%)

      Query: 6   RLTRCQTALSQLSCDALLITNLNIFYLTFGSGTNATVLISPKHRIFVTDSRYALIAKNT 65
                R+ + + + D+LLIT++ NIFYLTFGSGT TV ++ K IF+TDSRY+ +A+
      Sbjct: 2   RIEKLKVKMLTENIDSLITDMKNIFYLTFGSGTAGTVFLTQKRNIPTDSRYSEMARGL 61

20   Query: 66   VREFDIIISREPLAAILKIIRDALIAIGFETDISYHMYKHMVEVFEDYRLIEAPSVVEK 125
                ++ F+II +R+P++ + ++ +++ + FE + Y +K + + L + V +
      Sbjct: 62   IKNFEIIESTRDPISLLTELSASESVKNMAFEETVDYAFFKRLSKAATKLDLFSTSNFVLE 121

25   Query: 126  LRMIKD 131
                LR IKD
      Sbjct: 122  LRQIKD 127

```

There is also homology to SEQ ID 3568.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1150

A DNA sequence (GBSx1226) was identified in *S.agalactiae* <SEQ ID 3569> which encodes the amino acid sequence <SEQ ID 3570>. This protein is predicted to be aminopeptidase P (pepQ-2). Analysis of this protein sequence reveals the following:

```

35   Possible site: 44
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.2508(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

45   >GP:CAA70068 GB:Y08842 aminopeptidase P [Lactococcus lactis]
      Identities = 131/205 (63%), Positives = 163/205 (78%), Gaps = 3/205 (1%)

      Query: 2   LDFIKPDRITTELQVANFLDFRMRELGATGPSFDFIVASGYRSAMPHGVASQKTIQSGETL 61
                L FI+P RT E++VANFLDF+MR+L A+G SF+ IVASG RS++PHGVA+ K IQ G+ +
      Sbjct: 149  LRFIEPGRT-EIEVANFLDFKMRDLEASGISFETIVASGKRSSLPHGVATSKMIQFGDPV 207

50   Query: 62   TLDFGCIYQHYVSDMTRTIHHVTDQEREIYDIVLKSQAIIGNVKSGMKRCDYDYLAR 121
                T+DFGCIY+HY SDMTRTI +G V D+ R IY+ V K+N+A+I VK+GM YD + R
      Sbjct: 208  TIDFGCIYEHYASDMTRTIFVGSVDDKMRITIVETVRKANEALIKQVKAGMTYAQYDNIPR 267

55   Query: 122  QVIENSGYGNHFTHGIGHGMGLDVHEIPYFGKS--EGVIASGMVVTDEPGIYLDNKYGVR 179

```

-1289-

+VIE + +G +FTHGIGHG+GLDVHEIPYF +S E + SGMV+TDEPGIYL GVR
 Sbjct: 268 EVIEKADFGQYFTHGIGHGLGLDVHEIPYFNQSMTEQLRSGMVTDEPGIYLPFEGGVR 327

Query: 180 IEDDLLITETGCEVLTSAPKELIVL 204
 IEDDLL+TE GCEVLT APKELIV+
 Sbjct: 328 IEDDLLVTENGCEVLTKAPKELIVI 352

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3567> which encodes the amino acid sequence <SEQ ID 3568>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1450(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/203 (71%), Positives = 171/203 (83%)

Query: 2 LDFIKPDRITTELQVANFLDFRMRELGATGPSFDFIVASGYRSAMPHGVASQKTIQSGETL 61
 LDFIKP TTE +ANFLDFRMR+ GA+G SFD IVASGY SAMPHG AS K IQ+ E+L
 Sbjct: 168 LDFIKPGTTTTERDLANFLDFRMROYGASGTSFDIIVASGYLSAMPHGRASDKVIQNKESL 227

Query: 62 TLDFGCIYQHYSVDMTRTIHIGHVTDQEREIYDIVLKSNAIIGNVKSGMKRCDYDYLAR 121
 T+DFGCIY HYVSDMTRTIHIG VTD+EREIY +VL +N+A+I +GM D+D + R
 Sbjct: 228 TMDFGCIYNHYVSDMTRTIHIGQVTDEEREIYALVLAANKALIAKASAGMTYSDFDGIPR 287

Query: 122 QVIENSGYGNHFTHGIGHGMGLDVHEIPYFGKSEGVIASGMVVTDEPGIYLDNKYGVRIE 181
 Q+I +GYG+ FTHGIGHG+GLD+HE P+FGKSE ++ +GMVVTDEPGIYLDNKYGVRIE
 Sbjct: 288 QLITEAGYGSRFTHGIGHGIGLDIHENPFFGKSEQLLQAGMVVTDEPGIYLDNKYGVRIE 347

Query: 182 DDLITETGCEVLTSAPKELIVL 204
 DDL+IT+TGC+VLT APKELIVL
 Sbjct: 348 DDLVITKTGCQVLTAPKELIVL 370

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1151

A DNA sequence (GBSx1227) was identified in *S.agalactiae* <SEQ ID 3571> which encodes the amino acid sequence <SEQ ID 3572>. This protein is predicted to be yfhC protein (comEB). Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1401(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA05053 GB:AP001511 late competence operon required for DNA
 binding and uptake [Bacillus halodurans]
 Identities = 78/146 (53%), Positives = 107/146 (72%)

Query: 1 MNRLSWEDYFMANAELISKRSTCDRAFGAVLVKNNRIIATGYNGGVSETDNCNEVGHYM 60
 MNR+SW+ YFMA + L++ RSTC R VGA +V++ RIIA GYNG +S +C + G Y+
 Sbjct: 1 MNRISWDQYFMAQSHLLALRSTCTRLMVGATIVRDKRIIAGGYNGSISGGPHCIDEGCYV 60

-1290-

Query: 61 EDGHCIRTVHAEMNALIQCAKEGISTNNTETIYVTHFPCINCTKALLQAGVKKITYKANYR 120
 +GHCIRT+HAE+NAL+QCAK G+ T EIYVTHFPC+NCTKA++Q+G+KK+ Y +Y+
 Sbjct: 61 VEGHCIRTIHAEVNALLQCAKFGVPTGAEIYVTHFPCVNCTKAIQSGIKKVYYATDYK 120

Query: 121 HPFAIELMEAKGVAYVQHDVPEVTL 146
 P+A EL GV Q ++ E+ L
 Sbjct: 121 NSPYAEELFRDAGVDVEQVELEEMIL 146

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3573> which encodes the amino acid sequence <SEQ ID 3574>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3155(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 133/146 (91%), Positives = 140/146 (95%)

- Query: 2 NRLSWEDYFMANAELISKRSTCDRAFGAVLVKNNRIIATGYNGGVSETDNCNEVGHYME 61
 NRLSW+DYFMANAELISKRSTCDRAFGAVLVK+NRRIATGYNGGVS TDNCNE GHYME
 25 Sbjct: 18 NRLSWQDYFMANAELISKRSTCDRAFGAVLVKDNRIIATGYNGGVSATDNCNEAGHYME 77
 Query: 62 DGHHCIRTVHAEMNALIQCAKEGISTNNTETIYVTHFPCINCTKALLQAGVKKITYKANYRP 121
 DGHHCIRTVHAEMNALIQCAKEGIST+ TEIYVTHFPCINCTKALLQAG+ KITYKA+YRP
 30 Sbjct: 78 DGHHCIRTVHAEMNALIQCAKEGISTDGTETIYVTHFPCINCTKALLQAGITKITYKAHYRP 137
 Query: 122 HPFAIELMEAKGVAYVQHDVPEVTLG 147
 HPFAIELME KGVAYVQHDV++ LG
 Sbjct: 138 HPFAIELMEKKGVAYVQHDVPQIVLG 163

- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1152

A DNA sequence (GBSx1228) was identified in *S.agalactiae* <SEQ ID 3575> which encodes the amino acid sequence <SEQ ID 3576>. Analysis of this protein sequence reveals the following:

- 40 Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.2454(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1291-

Example 1153

A DNA sequence (GBSx1229) was identified in *S.agalactiae* <SEQ ID 3577> which encodes the amino acid sequence <SEQ ID 3578>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 25
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL      Likelihood = -1.65      Transmembrane      4 - 20 ( 3 - 21)

      ----- Final Results -----
10          bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1154

A DNA sequence (GBSx1230) was identified in *S.agalactiae* <SEQ ID 3579> which encodes the amino acid sequence <SEQ ID 3580>. Analysis of this protein sequence reveals the following:

```

20      Possible site: 54
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
25          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

30      >GP:BAB04699 GB:AP001510 unknown conserved protein [Bacillus halodurans]
          Identities = 47/94 (50%), Positives = 65/94 (69%)

      Query: 2  LLPVGSVVYLIDGNQKLIVIVNRGAIVEQEGQEVYFDYLG GIFPEGLNLEQVYYFNQEDID 61
          +LP+GS+VYL +G KL+I+NRG I+E G+ FDY G +P+GL ++V+YFN E+ID
      Sbjct: 1  MLPIGSIVYLKEGTSKLMILNRGPFILEANGENKMFDSGCFYPQGLVPDKVFYFNHENID 60

      Query: 62  EVVFEgyHDEEEERVSR LIEKWKNTegKNLPKGK 95
          EVVFEg+ D+EE+R +L WK KGK
      Sbjct: 61  EVVFEgFQDDEEQRFQKLFDWKKENKDRYVKGK 94

```

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1155

45 A DNA sequence (GBSx1231) was identified in *S.agalactiae* <SEQ ID 3581> which encodes the amino acid sequence <SEQ ID 3582>. Analysis of this protein sequence reveals the following:

```

      Possible site: 15
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50          bacterial cytoplasm --- Certainty=0.3560(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-1292-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1156

A DNA sequence (GBSx1232) was identified in *S.agalactiae* <SEQ ID 3583> which encodes the amino acid sequence <SEQ ID 3584>. This protein is predicted to be elongation factor p (efp). Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 15 bacterial cytoplasm --- Certainty=0.3067 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:CAB14376 GB:Z99116 elongation factor P [Bacillus subtilis]
Identities = 89/186 (47%), Positives = 120/186 (63%), Gaps = 1/186 (0%)
- Query: 1 MIEASKLKAGMTFETADGKLIRVLEASHHKPGKGNTIMRMKLRDVRTGSTFDTSYRPEEK 60
MI + + G+T + DG + RV++ H KPGKG +R KLR++RTG+ + ++R EK
25 Sbjct: 1 MISVNDFRITGLTIDV-DGGIWRVVDVQHVKPGKGAAFVRSKLRNLRGTGAIQEKTFRAGEK 59
- Query: 61 FEQAIIEIVPAQYLYKMDDTAYFMNNETYDQYEIPTVNIENELLYILENSEVVKIQFYGTE 120
+A IET QYLY D FM+ +Y+Q E+ IE EL Y+LEN V I Y E
30 Sbjct: 60 VAKAQIETKTMQYLYANGDQHVMDTSSYQLELSATQIEEELKYLLENMSVHIMMYQDE 119
- Query: 121 VIGVQIPTTVELTVAETQPSIKGATVTGSGKPATMETGLVVNVPDFIEAGQKLVINTAEG 180
+G+++P TVEL V ET+P IKG T +G KPA ETGLVVNV F+ G LV+NT++G
Sbjct: 120 TLGIELPNTVELKVVETEPIKGD TASGGTKPAKTETGLVVNVPPFVNEGDTLVVNTSDG 179
- 35 Query: 181 TYVSRA 186
+YVSRA
Sbjct: 180 SYVSRA 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3585> which encodes the amino acid sequence <SEQ ID 3586>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 45 bacterial cytoplasm --- Certainty=0.1813 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 50 Identities = 170/186 (91%), Positives = 180/186 (96%), Gaps = 1/186 (0%)
- Query: 1 MIEASKLKAGMTFETADGKLIRVLEASHHKPGKGNTIMRMKLRDVRTGSTFDTSYRPEEK 60
MIEASKLKAGMTFE A+GKLIRVLEASHHKPGKGNTIMRMKLRDVRTGSTFDT+YRP+EK
Sbjct: 1 MIEASKLKAGMTFE-AEGKLIRVLEASHHKPGKGNTIMRMKLRDVRTGSTFDTTYRPDEK 59
- 55 Query: 61 FEQAIIEIVPAQYLYKMDDTAYFMNNETYDQYEIPTVNIENELLYILENSEVVKIQFYGTE 120

-1293-

```

                FEQAIIEITVPAQYLYKMDDTAYFMN +TYDQYEIP N+E ELLYILENS+VKIQFYG+E
Sbjct: 60 FEQAIIEITVPAQYLYKMDDTAYFMNTDTYDQYEIPVANVEQEELLYILENSDVKIQFYGSE 119

Query: 121 VIGVQIPTTVELTVAETQPSIKGATVTGSGKPATMETGLVVNPVDFIEAGQKLVINTAEG 180
5          VIGV +PTTVELTVAETQPSIKGATVTGSGKPAT+ETGLVVNPVDFIEAGQKL+INTAEG
Sbjct: 120 VIGVTVPTTVELTVAETQPSIKGATVTGSGKPATLETGLVVNPVDFIEAGQKLIINTAEG 179

Query: 181 TYVSRA 186
          TYVSRA
10 Sbjct: 180 TYVSRA 185

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1157

15 A DNA sequence (GBSx1233) was identified in *S.galactiae* <SEQ ID 3587> which encodes the amino acid sequence <SEQ ID 3588>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1508(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

25 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06505 GB:AP001516 unknown conserved protein [Bacillus halodurans]
Identities = 42/107 (39%), Positives = 70/107 (65%), Gaps = 4/107 (3%)

Query: 5  NLGEIVISPRVLEVITGIAATKVDGVHSLRNK--AVTDSLKSKSLGRGVYLKNEEDDTV 61
30      +LG + ISP V+EVI GIAA++V+GV ++R      V + L K+ G+GV + + D+ +
Sbjct: 15 DLGRVEISPEVIEVIAGIAASEVEGVATMRGNFAAGVAEKLGYKNHKGKGVKV-DLNDEGI 73

Query: 62 AADIYVYLQYGVNVPVAVSIAIQQAVKTAVYDMAEVKISSVNIHVHVEGI 108
          D+ V + YGV+VP V+  IQQ +K A+  M  +++ S+N+H+ G+
35 Sbjct: 74 IVDVSVIILYGVSVPEVAKKIQONIKQALQMTAIELQSINVHIVGV 120

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3589> which encodes the amino acid sequence <SEQ ID 3590>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0882(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 101/129 (78%), Positives = 113/129 (87%)

50 Query: 1  MTTENLGEIVISPRVLEVITGIAATKVDGVHSLRNKAVTDSLKSKSLGRGVYLKNEEDDT 60
      MTE +GEIVISPRVLEVITGIA T+V+GVHSL NK + DS +K SLG+GVYL+ EED +
Sbjct: 1  MTTEYIGEIVISPRVLEVITGIATTQVEGVHSLHNKKMADSFNKASLGKGVYLQTEEDGS 60

Query: 61  VAADIYVYLQYGVNVPVAVSIAIQQAVKTAVYDMAEVKISSVNIHVHVEGIVPEKTPKPD LKS 120
55 V ADIYVYLQYGV VP VS+ IQ+ VK+AVYDMAEV IS+VNIHVHVEGIV EKTTPKPD LKS
Sbjct: 61 VTADIYVYLQYGVKVPTVSMNIQKTVKSAVYDMAEVPISAVNIHVHVEGIVAEKTPKPD LKS 120

Query: 121 LFDEDFLDD 129
          LFDEDFLDD

```

-1294-

Sbjct: 121 LFDEDFLDD 129

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1158

A DNA sequence (GBSx1234) was identified in *S.agalactiae* <SEQ ID 3591> which encodes the amino acid sequence <SEQ ID 3592>. This protein is predicted to be a utilization substance protein b homolog (nusB). Analysis of this protein sequence reveals the following:

Possible site: 27
 10 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.32 Transmembrane 48 - 64 (47 - 64)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
 15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14363 GB:Z99116 similar to transcription termination
 20 [Bacillus subtilis]
 Identities = 51/129 (39%), Positives = 82/129 (63%), Gaps = 9/129 (6%)
 Query: 9 RRDLRERAFQTLFSLETGGEFIDAAHFAYGYDKTVSEDKVLEVPFLLNLVNGVVDHKDE 68
 RR RE+A Q LF ++ ++ A + + E+K F LV+GV++H+D+
 25 Sbjct: 3 RRTAREKALQALFQIDVSDIADVNEA-----IEHALDEEKT---DPFFEQLVHGVLEHQDQ 54
 Query: 69 LDTLISSHLKSGWSLERLTLVDKSLRLGLYEIKYFDETPDRVALNEIIEIAKKYSDETS 128
 LD +IS HL + W L+R+ VD+++LRL YE+ Y ++ P V++NE IE+AK++ D+ +
 30 Sbjct: 55 LDEMISKHLVN-WKLDRIANVDRAILRLAAYEMAYAEDIPVNVSMNEAIEIAKRFGDDKA 113
 Query: 129 AKFVNGLLS 137
 KFVNG+LS
 Sbjct: 114 TKFVNGVLS 122

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3593> which encodes the amino acid sequence <SEQ ID 3594>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 40 INTEGRAL Likelihood = -1.75 Transmembrane 53 - 69 (53 - 69)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14363 GB:Z99116 similar to transcription termination
 [Bacillus subtilis]
 Identities = 47/134 (35%), Positives = 76/134 (56%), Gaps = 10/134 (7%)
 50 Query: 15 RRDLRERAFQALFNIEGAEELAAASQFAYGYDKVTGEDAQVLELPFLLSLVTGVNNHKE 74
 RR RE+A QALF I++ +++ + D+ + F LV GV H++
 Sbjct: 3 RRTAREKALQALFQIDV-SDIADVNEAIEHALDEEKTDP-----FFEQLVHGVLEHQD 53
 Query: 75 ELDNLISTHLKKGWSLERLTLTDKTLRLGLFEIKYFDKTPDRVALNEIIEVVKKYSDET 134
 +LD +IS HL W L+R+ D+ +LRL +E+ Y + P V++NE IE+ K++ D+
 55 Sbjct: 54 QLDDEMISKHLVN-WKLDRIANVDRAILRLAAYEMAYAEDIPVNVSMNEAIEIAKRFGDDK 112

-1295-

Query: 135 SAKFINGLLSQYVS 148
 + KF+NG+LS S
 Sbjct: 113 ATKFVNGVLSNIKS 126

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 104/142 (73%), Positives = 125/142 (87%), Gaps = 1/142 (0%)

Query: 1 MTSVFKDSRRDLRERAFQTLFSLETGGEFIDA AHFAYGYDKTVSED-KVLEVP IFLNLV 59
 MT+ F++SRRDLRERAFQ LF++E G E + A+ FAXGYDK ED +VLE+PIFLL+LV
 10 Sbjct: 7 MTNSFQNSRRDLRERAFQALFNIEMGAELLAASQFAYGYDKVTGEDAQVLELPIFLLSLV 66
 Query: 60 NGVVDHKDELDTLISSHLKSQWLERLTLVDKSLRLGLYEIKYFDETPDRVALNEIIEI 119
 GV +HK+ELD LIS+HLK GWSLERLTL DK+LLRLGL+EIKYFD+TPDRVALNEIIE+
 15 Sbjct: 67 TGVNNHKEELDNLISTHLKKGWSLERLTLTDKTLRLGLFEIKYFDKTPDRVALNEIIEV 126
 Query: 120 AKKYSDETS AKFVNGLLSQFIT 141
 KKYSDETS AKF+NGLLSQ+++
 Sbjct: 127 VKKYSDETS AKFINGLLSQYVS 148

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1159

A DNA sequence (GBSx1235) was identified in *S.agalactiae* <SEQ ID 3595> which encodes the amino acid sequence <SEQ ID 3596>. Analysis of this protein sequence reveals the following:

25 Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.81 Transmembrane 239 - 255 (239 - 255)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAC31628 GB:U46902 ScrR [Streptococcus mutans]
 Identities = 225/320 (70%), Positives = 273/320 (85%)
 Query: 1 MVAKLTDVAALAGVSPTTVSRVINKKGYLSQKTVTKVNEAMRTLGYKPNNLARS LQGKSA 60
 MVAKLTDVA LAGVSPTTVSRVIN+KGYLS+KT+TKV AM+TLGYKPNNLARS LQGKSA
 40 Sbjct: 1 MVAKLTDVAALAGVSPTTVSRVINRKGYLSEKTTITKVQAAMKTLGYKPNNLARS LQGKSA 60
 Query: 61 KLIGLIFPNIRNIFYAELIEHLEIELFKHGYKTILCNSEKDPIKEKEYLEMLGANQVDGI 120
 KLIGLIFPNI +IFY+ELIE+LEIELFKHGYK I+CNS+ +P KE++YLEML ANQVDGI
 45 Sbjct: 61 KLIGLIFPNISHIFYSELIEYLEIELFKHGYKAIICNSQNNPDKERDYLEMLEANQVDGI 120
 Query: 121 ISSSHNLGIDDYEKVEAPIVAFDRNLAPHIPIVSSDNFFGGKMAAQTLKKHGCQKMIMIT 180
 ISSSHNLGIDDYEKV API+AFDRNLAP+IPIVSSDNF GG+MAA+ LKKHGCQ IMI
 Sbjct: 121 ISSSHNLGIDDYEKVSAPIIAFDRNLAPNIPIVSSDNFEGGRMAAKLLKKHGCQHPIMIA 180
 50 Query: 181 GNDNSDSPTGLRRLGFSYESKESKVITVTNGLSNMRREMEKSIISTHKPDGIFTSDDLIT 240
 G DNS+SPT LR+LGF ++ + ++ LS +R+EME+K I+ KPDGIF SDD+T
 Sbjct: 181 GKDNSNSPTALRQLGFKSVFAQAPIFHLSGELSIIRKEMEIKVILQNEKPDGIFLSDDMT 240
 Query: 241 ALLVIKLISQLGLSIPEDIKVIYDGTSTFIQDYVPHLTTIKQPIREIAQLMVEILLAKIE 300
 A+L +K+ +QL ++IP ++K+IGYDGT F+++Y P+LTTI+QPI++IA L+V+ILL KI+
 55 Sbjct: 241 AILTMKIANQLNITIPHELKIIGYDGTTHFVENYYPYLTITRQPIKDIAHLVDILLKKID 300
 Query: 301 GQKTNKDYILPVSLIPGSSV 320
 Q KDYILPV L+ G SV
 60 Sbjct: 301 HQDIPKDYILPVGLLSGESV 320

-1296-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3597> which encodes the amino acid sequence <SEQ ID 3598>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC31628 GB:U46902 ScrR [Streptococcus mutans]
 Identities = 226/321 (70%), Positives = 269/321 (83%), Gaps = 1/321 (0%)

Query: 1 VVAKLTDVAALAGVSPPTVSRVINKKGYLSQKTVNKNKAMRELGYKPNNLARSLQGKST 60
 +VAKLTDVA LAGVSPPTVSRVIN+KGYLS+KT+ KV AM+ LGYKPNNLARSLQGKS
 Sbjct: 1 MVAKLTDVAKLAGVSPPTVSRVINRKGYLEKTITKVQAAMKTLGYKPNNLARSLQKSA 60

Query: 61 QLIGLIFPNISNIFYAELIEHLEIELFKQGYKTIICNSEHNPVKEREYLEMLAANQVDGI 120
 +LIGLIFPNIS+IFY+ELIE+LEIELFK GYK IICNS++NP KER+YLEML ANQVDGI
 Sbjct: 61 KLIGLIFPNISHIFYSELIEFYLEIELFKHGYKAIICNSQNNPDKERDYLEMLEANQVDGI 120

Query: 121 ISSSHNLGIEDYERVEAPIVAFDRNLAPNIPVISSDNFEGGKLAAQTLQKHGCQNIVMIT 180
 ISSSHNLGI+DYE+V API+AFDRNLAPNIP++SSDNFEGG++AA+ L+KHGCQ+ +MI
 Sbjct: 121 ISSSHNLGIDDYEKVSAPIIAFDRNLAPNIPVSSDNFEGGRMAAKLLKKHGCQHPIMIA 180

Query: 181 GNDNSDSPTGLRQLGFNYQLKRSAEIILKPNNLSPVRREMEIKSILATRKPDGLFVSDDL 240
 G DNS+SPT LRQLGF + A I L LS +R+EMEIK IL KPDG+F+SDD+
 Sbjct: 181 GKDNSNSPTALRQLGFK-SVFAQAPIFHLSGELSIRKEMEIKVILQNEKPDGIFLSDDM 239

Query: 241 TAILIMKVAQLHITIPEDMKVIGYDGTTFIQQYVPQLATIRQPIDEIAKLSVEILIKKI 300
 TAIL MK+A QL+ITIP ++K+IGYDGT F++ Y P L TIRQPI +IA L V+IL+KKI
 Sbjct: 240 TAILTMKIANQLNITIPHELIKIGYDGTDFVENYYPYLTTRQPIKDIAHLLVDILLKKI 299

Query: 301 KKEKTSKDYILPITLLPGASI 321
 + KDYILP+ LL G S+
 Sbjct: 300 DHQDIPKDYILPVGLLSGESV 320

An alignment of the GAS and GBS proteins is shown below.

Identities = 247/321 (76%), Positives = 293/321 (90%), Gaps = 1/321 (0%)

Query: 1 MVAKLTDVAALAGVSPPTVSRVINKKGYLSQKTVTKVNEAMRTLGYKPNNLARSLQKSA 60
 +VAKLTDVAALAGVSPPTVSRVINKKGYLSQKTV KVN+AMR LGYKPNNLARSLQGKS
 Sbjct: 1 VVAKLTDVAALAGVSPPTVSRVINKKGYLSQKTVNKNKAMRELGYKPNNLARSLQKST 60

Query: 61 KLIGLIFPNIRNIFYAELIEHLEIELFKHGYKTILCNSEKDPIKEKEYLEMLGANQVDGI 120
 +LIGLIFPNI NIFYAELIEHLEIELFK GYKTI+CNSE +P+KE+EYLEML ANQVDGI
 Sbjct: 61 QLIGLIFPNISNIFYAELIEHLEIELFKQGYKTIICNSEHNPVKEREYLEMLAANQVDGI 120

Query: 121 ISSSHNLGIDDYEKVEAPIVAFDRNLAPHIPIVSSDNFFGGKMAAQTLKKHGCQKMIMIT 180
 ISSSHNLGI+DYE+VEAPIVAFDRNLAP+IP++SSDNF GKG+AAQTL+KHGCQ ++MIT
 Sbjct: 121 ISSSHNLGIEDYERVEAPIVAFDRNLAPNIPVISSDNFEGGKLAAQTLQKHGCQNIVMIT 180

Query: 181 GNDNSDSPTGLRRLGFSYESKES-KVITVTNGLSNMRREMEIKSIISTHKPDGIFTSDDL 239
 GNDNSDSPTGLR+LGF+Y+ K S ++I + N LS +RREME+KSI++T KPDG+F SDDL
 Sbjct: 181 GNDNSDSPTGLRQLGFNYQLKRSAEIILKPNNLSPVRREMEIKSILATRKPDGLFVSDDL 240

Query: 240 TALLVIKLISQLGLSIPEDIKVGIDGTSFIQDYVPHLTTIKQPIREIAQLMVEILLAKI 299
 TA+L++K+ QL ++IPED+KVGIDGT+FIQ YVP L TI+QPI EIA+L VEIL+ KI
 Sbjct: 241 TAILIMKVAQLHITIPEDMKVIGYDGTTFIQQYVPQLATIRQPIDEIAKLSVEILIKKI 300

Query: 300 EGQKTNKDYILPVSLIPGSSV 320
 + +KT+KDYILP++L+PG+S+
 Sbjct: 301 KKEKTSKDYILPITLLPGASI 321

-1297-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1160

- 5 A DNA sequence (GBSx1236) was identified in *S.agalactiae* <SEQ ID 3599> which encodes the amino acid sequence <SEQ ID 3600>. This protein is predicted to be sucrose-6-phosphate hydrolase (cscA). Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4775(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA35872 GB:X51507 sucrose-6-phosphate hydrolase [Streptococcus mutans]
Identities = 303/479 (63%), Positives = 359/479 (74%), Gaps = 25/479 (5%)

Query: 1 MNLPTAIRYPYDEWTEEDKENIVKNSKSPWRATYHLEAKTGLLNDPNGFSYFNGKFHL 60
MNLPT IRYR Y +WTEE+ ++I NV+ SPW TYH+E KTGLLNDPNGFSYFNGKF+L
Sbjct: 1 MNLPTQIRYRRYQDWTEEEIKSIKTNVALSPWHTTYHIEPKTGLLNDPNGFSYFNGKFNL 60

Query: 61 FYQNWPFGAAGHLKQWVHTESDDLHVHFKETGIKLPDHVNDSHGAYSGSALAIIDKFLF 120
FYQNWPFGAAGHLK W+HTES+DLVHFKETG L PD +DSHGAYSGSA I D+LFLF
Sbjct: 61 FYQNWPFGAAGHLKSWIHTESDDLHVHFKETGTVLYPDTSHDSHGAYSGSAYEIGDQLFLF 120

Query: 121 YTGNVDRMKWNRDPRQIGAWMTNDGKITKFDKVLISQPNVDTEHFRDPQIFNYDNQFYAV 180
YTGNVDR W R P QIGA+M G I KF VLI QPNVDTEHFRDPQIFNY QFYA+
Sbjct: 121 YTGNVDRDENWVRHPLQIGAFMDKKGNIQKFTDVLKQPNVDTEHFRDPQIFNYKQFYAI 180

Query: 181 IGAQNSKKCGFIKLYKALNNDIHHWEFVGDLDGFGGTSEYMIECPNIIIFVKGKPVLLYSP 240
+GAQ+ LDFGG+ SEYMIECPN++F+ +PVL+YSP
Sbjct: 181 VGAQS-----LDFGGSKSEYMIECPNLVFINEQPVLIIYSP 215

Query: 241 QGLDKNELDYQNIYPNTYKIGQYFDANSSKIVEPSPIYNLDYGFAYATQGFNTSDGRAF 300
QGL K+ELDY NIYPNTYK+ Q FD +V+ S I NLD+GFE YATQ FN DGR +
Sbjct: 216 QGLSKSELDYHNIYPNTYKVCQSFDEKPALVDASEIQNLDFGFECYATQAFNAPDGRVY 275

Query: 301 IVSWIGLPDIDYPSDQFDYQGAMSLVKELSIKNGNLYQYVPVAMKNLRQHQAEFKTQLQT 360
VSWIGLPDIDYPSD +DYQGA+SLVKELS+K+G LYQYVPV A+++LR + +T
Sbjct: 276 AVSWIGLPDIDYPSDSYDQGALSILVKELSLKHGKLYQYVPEAVRSLRSEKEAVTYKPE 335

Query: 361 NNTYELELLVPRNDLSSSVFLFANPKGQGLSITITDVKGVIIIDRSQAGQYATEFGTSRQ 420
NNTYELEL + ++ +LFA+ KG GL+IT+DT G ++IDRS+AG+QYA EFG+ R
Sbjct: 336 NNTYELELTFDSSSVNELLLFADNKGNGLAITVDTKMGITILIDRSKAGEQYALEFGSQRS 395

Query: 421 CDIPKDATSINIFIDKSI FEIFINKGEKVFTGRVFPDAEQSGIQLKEGHVHGKYFELKY 479
C I T +NIF+DKSIFEIFINKGEKVFTGRVFP+ +Q+GI +K G G Y+ELKY
Sbjct: 396 CSIQAKETVVNIFVDKSI FEIFINKGEKVFTGRVFPNDKQTGIVIKSGKPSGNYELKY 454

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3601> which encodes the amino acid sequence <SEQ ID 3602>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4629(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1298-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/479 (60%), Positives = 367/479 (76%)

```

5      Query: 1  MNLPTIIRYPYDEWTEEDKENIVKNVSKSPWRATYHLEAKTGLLNDPNGFSYFNGKFHL 60
      M+LP IRYRYPY EW+ +D + I + +++SPW + +H+E KTGLLNDPNGFSYFNG++HL
      Sbjct: 2  MDLPQAIIRYPYKEWSSKDYQAITEKMAQSPWHSQFHVEPKTGLLNDPNGFSYFNGRYHL 61

10     Query: 61  FYQNWPFGAAHGLKQWVHTESDDLHVHFETGKIKLKPDPVNDSHGAYSGSALAIDDKLFLF 120
      FYQNWP+GAAHGLKQWVH S DLVHF ET +L PDH +DSHGAYSGSA AIDDKLFLF
      Sbjct: 62  FYQNWPYGAHGLKQWVHMTSTDLVHFTETRSRLLPDHAHDSHGAYSGSAYAIIDDKLFLF 121

15     Query: 121 YTGNVDRMKWNRDPQIGAWMTNDGKITKFDKVLISQPNVTEHFRDPQIFNYDNQFYAV 180
      YTGNVDR W R P Q+GAWM G I+K +VLI QP+DVTEHFRDPQ+F+Y QFYA+
      Sbjct: 122 YTGNVDRANWVRTPLQVGAWMDKQGNISKIPQVLIQPDVTEHFRDPQLFSYQGQFYAI 181

20     Query: 181 IGAQNSKKCGFIKLYKALNNDIHHWEFVGDLDFGGTGSEYMIECPNIIFVKGKPVLLYSP 240
      IGAQ G IKLYKA++N + +W F+ DLDF +G+EYMIECPN++FV KPVL++SP
      Sbjct: 182 IGAQGLDGKGIKLYKAVDNHVDNWRFIADLDFDSDGTEYMIECPNLVVFDDKPVLI FSP 241

25     Query: 241 QGLDKNELDYQNIYPNTYKIGQYFDANSSKIVEPSPIYNLDYGFAYATQGFNTSDGRA 300
      QGL K +LDYQNIYPNTYKI + F+ + +++ + NLD+GFAYATQ F++ DGR
      Sbjct: 242 QGLAKADLDYQNIYPNTYKIFESFNPETGQLLGGGALQNLDFGFAYATQAFSSPDGRVL 301

30     Query: 301 IVSWIGLPDIDYPSDQFDYQGAMSLVKELSIKNGNLYQYVPVAMKNLRQHQAEFKTQLQT 360
      VSWIGLPDIDYP+D++DYQGA+SLVKEL IK+G LYQ PV A++NLR F ++ +
      Sbjct: 302 AVSWIGLPDIDYPTDRYDYQGALSLVKELRIKDGIYQTPVSALQNLRGPAELFHNKIDS 361

35     Query: 361 NNTYELELLVPRNDLSSFVLFAFPKGQGLSITIDTVKGKVIIDRSQAGQQYATEFGTSRQ 420
      +N YELEL +P +LFA+ KG GL + +DT KG++ IDRS+AG QYA ++GT R
      Sbjct: 362 SNCYELELTIPGQKLDLLLFADQKGNLRLKVDTTKGQLSIDRSRAGVQYAQDYGTVRS 421

40     Query: 421 CDIPKDATSINIFIDKSIFEIFINKGEKVFTGRVFPDAEQSGIQLKEGHVHGKYFELKY 479
      C IP+ ++N+++D SI EIFIN+G+KV T RVFP Q+GIQ+ EG G Y+E++Y
      Sbjct: 422 CQIPQGHVTLNVYVDNSILEIFINQGQKVLTSRVFPTHGQTGIQVVEGQAFGHYEMRY 480

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1161

A DNA sequence (GBSx1237) was identified in *S.agalactiae* <SEQ ID 3603> which encodes the amino acid sequence <SEQ ID 3604>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

      bacterial cytoplasm --- Certainty=0.2204(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1299-

Example 1162

A DNA sequence (GBSx1238) was identified in *S.agalactiae* <SEQ ID 3605> which encodes the amino acid sequence <SEQ ID 3606>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -7.64	Transmembrane	259 - 275 (250 - 283)
INTEGRAL	Likelihood = -4.41	Transmembrane	113 - 129 (109 - 130)
INTEGRAL	Likelihood = -3.03	Transmembrane	180 - 196 (180 - 196)
INTEGRAL	Likelihood = -3.03	Transmembrane	439 - 455 (438 - 456)
INTEGRAL	Likelihood = -2.81	Transmembrane	298 - 314 (298 - 317)
INTEGRAL	Likelihood = -2.02	Transmembrane	396 - 412 (395 - 412)

----- Final Results -----

bacterial membrane --- Certainty=0.4057(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC99320 GB:AF059741 sucrose-specific PTS permease [Clostridium
beijerinckii]

Identities = 235/453 (51%), Positives = 312/453 (67%), Gaps = 15/453 (3%)

Query: 7 IAKQVINAIGGASNVRSVAHCATRLRVMVKDETVIDKNTVENIEKVQGAFFNSGQYQIIF 66

+AK+++ IGG N++SV HCATRLR+++ D+ I++ +ENI+ V+G FF++ QYQII

Sbjct: 6 VAKEILENIGGKENIKSVEHCATRLRLILNDKEKINEKAIRENIDGVKGQFFSAAQYQIIL 65

Query: 67 GTGTVNKIYDEVVAQGLPTSSTSDQKAEAAKQGNFQRAIRTFGDVFPVPLPAIVATGLF 126

GTG VN++YD +V Q T + K EA Q Q+ RTFGDVFPV++P +VATGLF

Sbjct: 66 GTGFVNEVDVIVGQNSDLV-TGNNKEEAYSQMTLIQKISRFGDVFPVPIIPVLVATGLF 124

Query: 127 MGIRGAINNDTVLALFGTTSKAFSSNFYTYTVVLTDTAFAPFALISWSAFRVFGGNPV 186

MG+RG + N V + NF +T VLTDTAFAP PAL++WS + FGG PV

Sbjct: 125 MGLRGLLTNLGVQM-----NENFVLTQVLTDTAFAPLPAVWSTMKKFGGTPV 174

Query: 187 IGLVLGLMMVNSALPNAAVAVASGDAHPKIF--FGF-IPVVGYNVLPFAFFVGLLGAKLE 243

IG+V+GLM+V+ +LPNA+AVA+G A PI G IPVVGYNVLPFAFFVGLLGAKLE 243

Sbjct: 175 IGIIVIGLMLVSPSLPNAYAVAAGTATPINLTILGLNIPVVGYNVLPFAFFVGLLGAKLE 243

Query: 244 KWLHKKIPDVLDDLVPFLTFTVMSILALFVIGPIFHSVENYVLAGTKFVLNPLGLSGL 303

K L K +PDVLDL++ PF+T +L L ++GPI H+ E + K + LP GL GL

Sbjct: 235 KALKKVVDPVLDLIVTFPITLFLSMVLGLLIVGPIMHNAEQIFGAIKGFMGLPFGGL 294

Query: 304 ILGGVHQIIVVTGVHHIFNLLEAQLIAADGKDPFNAIITAAMTAQAGATLAVGVKTKNKK 363

++GGVHQ+IVVTGVHH N LE +L+++ GKD FNA+IT + AQ A LAV VKTK+KK

Sbjct: 295 VVGGVHQLIIVVTGVHHALNALEVELLSSTGKDAFNAMITCGIVAQGAALAVAVKTKDKK 354

Query: 364 LKALAFPAALSAGLGITEPAIFGVNLRFGKPFIMGLIAGAAGGWLASILKLAGTGFGITI 423

++L +A+ A LGITEPAIFGVNLRFGKPFIMGLIAGAAGGWLASILKLAGTGFGITI 423

Sbjct: 355 KRSLYISSAIPAFGLGITEPAIFGVNLRFGKPFIMGLIAGAAGGWLASILKLAGTGMGITA 414

Query: 424 IPGTLLEYLNGQIVKYLMVIGTTSFVLTVMF 456

+PG LLY+N + Y+++ + ++AF LT F

Sbjct: 415 LPGMLLYVN-NLGSYILVNVVAIAVAFCLTLFF 446

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3607> which encodes the amino acid sequence <SEQ ID 3608>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -4.99	Transmembrane	111 - 127 (108 - 129)
INTEGRAL	Likelihood = -4.57	Transmembrane	176 - 192 (176 - 193)
INTEGRAL	Likelihood = -4.35	Transmembrane	436 - 452 (431 - 453)
INTEGRAL	Likelihood = -3.88	Transmembrane	295 - 311 (293 - 314)
INTEGRAL	Likelihood = -3.50	Transmembrane	259 - 275 (253 - 277)

-1300-

INTEGRAL Likelihood = -2.07 Transmembrane 405 - 421 (405 - 421)
 INTEGRAL Likelihood = -0.43 Transmembrane 219 - 235 (219 - 235)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAC99320 GB:AF059741 sucrose-specific PTS permease [Clostridium
 beijerinckii]
 Identities = 234/451 (51%), Positives = 312/451 (68%), Gaps = 11/451 (2%)

15 Query: 1 MDNRQIAAEVIEALGGRENVRVAHCATRLRVMVYDEGKIDKEKAEIDKVGGAFFNSGQ 60
 M + +A E++E +GG+EN++SV HCATRLR+++ D+ KI+++ E ID VKG FF++ Q
 Sbjct: 1 MKEQIVAKBILENIGGKENIKSVEHCATRLRLILNDKEKINEKAIEINIDGVKGQFFSAAQ 60

20 Query: 61 YQMIFGTGTVNNIYDEVVALGLPTSSTSEQKAEAGKHGNIFQRAIRTFGDVFVPIIPAIV 120
 YQ+I GTG VN +YD +V T K EA + Q+ RTFGDVFVPIIP +V
 Sbjct: 61 YQIILGTGFVNEVDVIVGQNSDLV-TGNNKEEAYSQMTLIQKISRTFGDVFVPIIPVLV 119

25 Query: 121 ATGLFMGVRGLVTQPAIMDLFGVHEYGENFLMYTRILTDTAFFVYLPALVAWSAFRVFGG 180
 ATGLFMG+RGL+T + + ENF+++T++LTDTAFF +LPALVAWS + FGG
 Sbjct: 120 ATGLFMGLRGLLTNLGV-----QMNFVFLFTQVLTDTAFAFLPALVAWSMTKKFGGT 172

30 Query: 181 PIIGIVLGLMLVSNELPNNAVVASGGDVK-PLTFFGF-VPVVGYYQGTVLPFAFFVGLVGAK 238
 P+IGIV+GLMLVS LPNA+ VA+G LT G +PVVGYQG+VLPA +G++ AK
 Sbjct: 173 PVIGIVLGLMLVSPSLPNAYAVAAGTATPINLTILGLNIPVVGYYQGSVLPALVGLGIAAK 232

35 Query: 239 LEKWLHKKVPEALDLLVTPFLTFAIMSTLGLFVIGPVFHSLENLVLAGTQAVLHLPFGIA 298
 +K L K VP+ LDL+VTFP+T LGL ++GP+ H+ E L+ + + LPPG+
 Sbjct: 233 TQKALKKVVDPDVLDLIVTPFITLLFSMVLGLLIVGPIMHNAEQLIFGAIKGFMGLPFGLG 292

40 Query: 299 GLIVGGIQQQLIVVTGIHHIFNFLEAQLIANTGKDPFNAYLTAATAAQAGATLAVAVKTKS 358
 GL+VGG+ QLIVVTG+HH N LE +L+++TGKD FNA +T AQ A LAVAVKTK
 Sbjct: 293 GLVVGGVHQLIVVTGVHHALEVEVLLSSTGKDAFNAMITCGIVAQGAALAVAVKTKD 352

45 Query: 359 TKLKGLAFTPSTLSALLGITEPAIFGVNLRYPKVFVSLIGGALGGWVAGLFGIAGTGFGI 418
 K + L S + A LGITEPAIFGVNLR+ K F+ G GGA+GG ++G+ +AGTG GI
 Sbjct: 353 KKKRSLYISSAIPAFLGITEPAIFGVNLRFIKPFIFGCAGGAVGGMLSGILHLAGTGMGI 412

Query: 419 TVLPGTLLYLNQQLLYLVTMLVGLGVAF 449
 T LPG LLY+N L Y++ +V + VAF +
 Sbjct: 413 TALPGMLLYVN-NLGSYILVNVVAIAVAFCL 442

An alignment of the GAS and GBS proteins is shown below.

Identities = 409/618 (66%), Positives = 491/618 (79%), Gaps = 12/618 (1%)

50 Query: 4 NTEIAKQVINAIGGASNVRSVAHCATRLRVMVKDETVIDKNTVENIEKVQGAFFNSGQYQ 63
 N +IA +VI A+GG NVRSVAHCATRLRVMV DE IDK E I+KV+GAFFNSGQYQ
 Sbjct: 3 NRQIAAEVIEALGGRENVRVAHCATRLRVMVYDEGKIDKEKAEIDKVGGAFFNSGQYQ 62

55 Query: 64 IIFGTGTVNKIYDEVVAQGLPTSSTSDQKAEAAKQGNFQRAIRTFGDVFVPLLPVPAIVAT 123
 +IFGTGTVN IYDEVVA GLPTSSTS+QKAE K GN FQRAIRTFGDVFVPP++PAIVAT
 Sbjct: 63 MIFGTGTVNNIYDEVVALGLPTSSTSEQKAEAGKHGNIFQRAIRTFGDVFVPIIPAIVAT 122

60 Query: 124 GLFMGIRGAINNDTVLALFGTTSKAFSSSNFYTYTVLTDTAFAFFPALISWSAFRVFGG 183
 GLFMG+RG + ++ LPG NF YT +LTDTAFF + PAL++WSAFRVFGG
 Sbjct: 123 GLFMGVRGLVTQPAIMDLFGVHEYG---ENFLMYTRILTDTAFFVYLPALVAWSAFRVFGG 179

65 Query: 184 NPVIGLVGLMMVNSALPNAAVASG-DAHPIKFFGFIPVVGYYQNSVLPFAFFVGLLGAKL 242
 NP+IG+VLGLM+V++ LPNAW VASG D P+ FFGF+PVVGYQ +VLPFAFFVGL+GAKL
 Sbjct: 180 NPIIGIVLGLMLVSNELPNNAVVASGGDVKPLTFFGFVPVVGYYQGTVLPFAFFVGLVGAKL 239

Query: 243 EKWLHKKIPDVLDDLVPFLTFVMSILALFVIGPIFHSVENYVLAGTKFVLNPLGLSG 302
 EKWLHKK+P+ LDLL+ PFLTF +MS L LFVIGP+FHS+EN VLAGT+ VL+LP G++G

-1301-

Sbjct: 240 EKWLHKKVPEALDLLVTPFLTFAIMSTLGLFVIGPVFHSLENLVLAGTQAVLHLPFGIAG 299

Query: 303 LILGGVHQIIVVTGVHHIFNLLEAQLIAADGKDPFNAIITAAMTAQAGATLAVGVKTKNK 362
LI+GG+ Q+IVVTG+HHIFN LEAQLIA GKDPFNA +TAA AQAGATLAV VKTK+

5 Sbjct: 300 LIVGGIQQLIVVTGIIHHIFNLEAQLIANTGKDPFNAYLTAATAAQAGATLAVAVKTKST 359

Query: 363 KLKALAFPAALSAGLGITEPAIFGVNLRFGKPFIMGLIAGAAGGWLASILKLAGTGFGIT 422
KLK LAFP+ LSA LGITEPAIFGVNLR+ K F+ GLI GA GGW+A + +AGTGFGIT

10 Sbjct: 360 KLKGLAFPSTLSALLGITEPAIFGVNLRYPKVFVSGLIGGALGGWVAGLFGIAGTGFGIT 419

Query: 423 IIPGTLLEYLNGQIVKYLMVIGTTSLAFVLTYMFGYEDKDEKAVAEVSPLVEETDDDDPTI 482
++PGTLLEYLNGQ+++YL+ ++ +AF + Y +GY+D++ + V V++T D P +

Sbjct: 420 VLPGTLLEYLNGQLLQYLVITMLVGLGVAFAIAYTWGYQDRETLPLPAVE--VDQTADQPAL 477

15 Query: 483 TQTSQLRAETIVSPLDGQVIALDVTSDPVFSSGIMGDGLAIKPRGNTIYSPVDGFGVQIAF 542
+ ET+ SPL+G V+ L VSDPVFSSG MG GLAIKP NT+YSPVDG V+I F

Sbjct: 478 AE-----ETLYSPLNGTVVDLSAVSDPVFSSGAMGQGLAIKPEDNTLYSPVDGKVEIVF 531

Query: 543 ETGHAYGIKSDKGABILIHIGIDVTMNGTGFTSKVKADQKVKKGDILGTFDSAKIAEAG 602
ETGHAY I S +GAE+L+HIGIDT +M G GF S V Q VKKGD+LG FD +KIAEAG

20 Sbjct: 532 ETGHAYAITSSQGAEVLLHIGIDTESMAGDGFESLVAVGQAVKKGDLLGHFDPKIAEAG 591

Query: 603 LDNTAMIIVTNTADFADV 620
LD+T M+IV+N AD+ V

25 Sbjct: 592 LDDTTMMIVSNIADYQSV 609

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1163

30 A DNA sequence (GBSx1239) was identified in *S.agalactiae* <SEQ ID 3609> which encodes the amino acid sequence <SEQ ID 3610>. This protein is predicted to be fructokinase. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2436(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA02467 GB:D13175 fructokinase [Streptococcus mutans]

Identities = 232/291 (79%), Positives = 257/291 (87%)

45 Query: 1 MTKLYGSIEAGGTFKVCVAGDEELKVVEKMQFPITTPQETIKKTVDFFKRFEKKLEAVAI 60
M+KLYGSIEAGGTFKVCVAGDE +++EK+QFPITTP ETI+KTV FFK+FE L +VAI

Sbjct: 1 MSKLYGSIEAGGTFKVCVAGDENFQILEKVQFPITTPYETIEKTVAFFKKFEADLASVAI 60

Query: 61 GSFPGPIDIDKKSPTYGYITTTPKLHWANVDLLGLISKDFNVFPYFTTVDVNSSAYGEVIAR 120
GSFPGPIDID+ S TYGYIT+TPK +WANVD +GLISKDF +PFYFTTVDVNSSAYGE IAR

50 Sbjct: 61 GSFPGPIDIDQNSDPTYGYITSTPKPNWANVDVFLGLISKDFKIPFYFTTVDVNSSAYGETIAR 120

Query: 121 NNIDSLVYYTIGTGIGAGAIQKGEFIGTGHTTEAGHTYMAHPQDQANDFKGICPFHNSC 180
+N+ SLVYYTIGTGIGAGAIQ GEFIGG GHTEAGH YMA HP D + F G CPFH C

55 Sbjct: 121 SNVKS LVYYTIGTGIGAGAIQNGEFIGGMGHTTEAGHVYMAPHPNDVHHGFVGTCPFHKGK 180

Query: 181 LEGLASGPTLEARTGIRGELIEENSMVWDVQAYYIAQAAIQATVLYRPQVIVFGGGVMAQ 240
LEGLA+GP+LEARTGIRGELIE+NS VWD+QAYYIAQAAIQATVLYRPQVIVFGGGVMAQ

60 Sbjct: 181 LEGLAAGPSLEARTGIRGELIEQNSEVWDIQAYYIAQAAIQATVLYRPQVIVFGGGVMAQ 240

Query: 241 EHMLRRVRQTFFATLLNGYLPVPDLSDYIVTPAIEENGSA TLGNFALAKKIS 291

-1302-

EHML RVR+ F +LLN YLPVPD+ DYIVTPA+ ENGSATLGN ALAKKI+
 Sbjct: 241 EHMLNRVREKFTSLNDYLPVPDVKDYIVTPAVAENGSATLGNLALAKKIA 291

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3611> which encodes the amino acid
 5 sequence <SEQ ID 3612>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2012(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 212/293 (72%), Positives = 246/293 (83%)

Query: 1 MTKLYGSIEAGGTFKVCavgDEELKVVEKMQFPTTTPQETIKKTVDFFKRFEKKLEAVAI 60
 M KLYGSIEAGGTFKVCavgDEE VV+K QFPTTTP+ETI +T+ +FK FE L +AI
 Sbjct: 1 MGKLYGSIEAGGTFKVCavgDEEFTVVDKTQFPTTTPETIARTIAYFKAFEADLAGMAI 60

20 Query: 61 GSFGPIDIDKKSCTYGYITTTTPKLNHWANVDLLGLISKDFNVPFYFTTDVNSSAYGEVIAR 120
 GSFGPIDID S+TYGYITTTTPK WANVDLLG +S F +PF TTDVNSSAYGEV+AR
 Sbjct: 61 GSFGPIDIDPSETYGYITTTTPKSGWANVDLLGQLSAAFKIPFDVTTDNSSAYGEVLAR 120

25 Query: 121 NNIDSLVYYTIGTGIGAGAIQKGEFIGGTGHTGHTYAMMHPDQANDFKGICPFHNSC 180
 ++SLVYYTIGTGIGAGAIQ G FIGG GHTGHTY+ HP D A F G+CPFH C
 Sbjct: 121 PGVESLVYYTIGTGIGAGAIQGHGFIGGLGHTGHTYVMPHPDDMAKGFLGVC PFHKG 180

30 Query: 181 LEGLASGPTLEARTGIRGELIEENSMVWDVQAYYIAQAAIQATVLYRPQVIVFGGGVMAQ 240
 LEG+A+GP++EARTG+RGE +++ + VWD+QA+YIAQAA+QAT+LYRPQVIVFGGGVMAQ
 Sbjct: 181 LEGMAAGPSIEARTGVRGERLDQEADVWDIQA FYIAQAALQATMLYRPQVIVFGGGVMAQ 240

35 Query: 241 EHMLRRVRQTFATLLNGYLPVPDLSDYIVTPAIEENGSATLGNFALAKKISK 293
 EHM+ RV F LL+GYLPVPDL+DYIVTPA+ +NGSATLGNFALAK ++G
 Sbjct: 241 EHMVLRVHDKFTALLSGYLPVPDLTDYIVTPAVADNGSATLGNFALAKLAAQG 293

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1164

40 A DNA sequence (GBSx1240) was identified in *S.agalactiae* <SEQ ID 3613> which encodes the amino acid sequence <SEQ ID 3614>. This protein is predicted to be Mannosephosphate Isomerase (pmi). Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.4717(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA04021 GB:D16594 Mannosephosphate Isomerase [Streptococcus mutans]
 Identities = 232/312 (74%), Positives = 262/312 (83%)

55 Query: 1 MSEPLFLEASMHDKIWGGTKLRDEFYDIPSETTGEYWAISAHPNGVSRVKNRFGKGCFL 60
 M PLFL++ MH KIWGG +LR EFGYDIPSETTGEYWAISAHPNGVS VKNG +KG L
 Sbjct: 1 MEGPLFLQSOMHKKIWGGNRLRKEFGYDIPSETTGEYWAISAHPNGVSVVKNGVYKGVPL 60

-1303-

Query: 61 DKLYQGEKSLFGNPDDTVFPPLLT KILDANDWLSVQVHPDDAYALKHEGELGKTECWYIIS 120
 D+LY + LFGN +VFPLLT KILDANDWLSVQVHPD+AYAL+HEGELGKTECWY+IS
 Sbjct: 61 DELYAHEHRELFNGSKSSVFPLLT KILDANDWLSVQVHPDNAYALEHEGELGKTECWYVIS 120

5 Query: 121 ADEGSEIIYGHNAKTKEELRQMIESGDWEHL LTRIPVKSGDFYVPSGTMHAIGKILIL 180
 ADEG+EIYGH AK+KEELRQMI +GDW+HLLT+IPVK+GDF+YVPSGTMHAIG+GI+IL
 Sbjct: 121 ADEGAEEIYGHAEKSKEELRQMIAAGDWDHLLTKIPVKAGDFFYVPSGTMHAIGRGIMIL 180

10 Query: 181 ETQQSSD TTYRVYDFDRPDASGKLRDLHIEQSIDVLTIGKPANTVPANMKL KHLSTLLV 240
 ETQQSSD TTYRVYDFDR D G+ R LHIEQSIDVLTIGKPAN PA + L+ L +T+LV
 Sbjct: 181 ETQQSSD TTYRVYDFDRKDDQGRKRALHIEQSIDVLTIGKPANATPAWLSLQGLETTVLV 240

15 Query: 241 SNDFFTVYKWEISGVTNFKQFAPYLLVSVLDGAGHITVDNKVYTLKKGDFILPNDVVKW 300
 S+ FFTVYKW+ISG +Q APYLLVSVL G G ITV + Y L+KGDH ILPN + W
 Sbjct: 241 SSPFFT VYKWIQISGSVKMQQTAPYLLVSVLAGQGRITVGLEQYALRKGDHILPNTIKSW 300

20 Query: 301 DIDGQLEIIASH 312
 DG LEIIASH
 Sbjct: 301 QFDGDLEIIASH 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3615> which encodes the amino acid sequence <SEQ ID 3616>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3714(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 232/312 (74%), Positives = 264/312 (84%)

35 Query: 1 MSEPLFLEASMHDKIWGGTKLRDEFGYDIPSETTG EYWAI SAHPNGVSRVKNRFRKGCFL 60
 MSEPLFL+++MHD+IWGGTKLRD F Y+IPS+TTGEYWAI SAHPNGVS V NGR++G L
 Sbjct: 1 MSEPLFLKSTMHDRIWGGTKLRDVFAYNIPSDTTGEYWAI SAHPNGVSTVTNGRYQGQPL 60

40 Query: 61 DKLYQGEKSLFGNPDDTVFPPLLT KILDANDWLSVQVHPDDAYALKHEGELGKTECWYIIS 120
 + LY E +LFGN + VFPLLT KILDANDWLSVQVHPDDAY +HEGELGKTECWYIIS
 Sbjct: 61 NTLYAQEPALFGNPKEEVFPPLLT KILDANDWLSVQVHPDDAYGREHEGELGKTECWYIIS 120

45 Query: 121 ADEGSEIIYGHNAKTKEELRQMIESGDWEHL LTRIPVKSGDFYVPSGTMHAIGKILIL 180
 A+EGSEI+YGH AK+KE+LR MIE+G W+ LLTR+PVK+GDF+YVPSGTMHAIGKILIL
 Sbjct: 121 AEEGSEIVYGHQAQSKEDLRAMIEAGAWDDLLTRVPVKAGDFFYVPSGTMHAIGKILIL 180

50 Query: 181 ETQQSSD TTYRVYDFDRPDASGKLRDLHIEQSIDVLTIGKPANTVPANMKL KHLSTLLV 240
 ETQQSSD TTYRVYDFDR D +G LRDHIE+SIDVLTIGKP N+VPA M L ++ +T LV
 Sbjct: 181 ETQQSSD TTYRVYDFDRKDVNGNLRDLHIEKSIDVLTIGKPENSVPATMVLDNMVATTLV 240

55 Query: 241 SNDFFTVYKWEISGVTNFKQFAPYLLVSVLDGAGHITVDNKVYTLKKGDFILPNDVVKW 300
 S FFTVYKW S + + KQ APYLLVSVL G G + VD K Y L+KG HFILPNDV W
 Sbjct: 241 STPFFT VYKWT SQMVDKQAAPYLLVSVLKGQGLYVDQKAYELEKGMHFILPNDVKS W 300

Query: 301 DIDGQLEIIASH 312
 DGQLE+I SH
 Sbjct: 301 SFDGQLEMIIVSH 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1304-

Example 1165

A DNA sequence (GBSx1241) was identified in *S.agalactiae* <SEQ ID 3617> which encodes the amino acid sequence <SEQ ID 3618>. This protein is predicted to be preprotein translocase *seca* subunit (*secA*).

Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.1102(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10107> which encodes amino acid sequence <SEQ ID 10108> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAA50286 GB:L32090 secA [Listeria monocytogenes]
    Identities = 503/843 (59%), Positives = 643/843 (75%), Gaps = 16/843 (1%)

20   Query: 11  MANILRTVIENDKGELKKLDKIAKKVDSYADHMAALSDEALQAKTPEFKERYQNGETLDQ 70
        MA +L+ + E+ K ++K L++ A ++ + AD AALSD+AL+ KT EFKER Q GETLD
    Sbjct: 1    MAGLLKKIFESGKKDKVLYLERKADEIIALADETAALSDDALREKTVEFKERVQKGETLDD 60

    Query: 71  LLPEAFVAVREASKRVLGLYPYHVQIMGGIVLHHGDIPEMRTGEGKTLTATMPVYLNAIS 130
        LL EAFAV RE +KR LGLYP+ VQ+MGGIVLH +I EM+TGEKTLTAT+PVYLNA+S
25   Sbjct: 61  LLVEAFAVAREGAKRALGLYPFKVQLMGGIVLHEDNIAEMKTGEGKTLTATLPVYLNALS 120

    Query: 131 GLGVHVTITVNEYLSTRDATEMGEVYSWLGLSVGINLAAKSPFEKREAYNCDITYSTNAEV 190
        G GVHV+TVNEYL+ RDA EMG +Y++LGLSVG+NL A S EKREAY CDITYSTN E+
30   Sbjct: 121 GEGVHVTVTNEYLAHRDAEEMGVLYNPLGLSVGLNLNALSSTEKREAYACDITYSTNNEL 180

    Query: 191 GFDYLDRDNMVRQEDMVQRPLNYALVDEVDVSLIDEARTPLIVSGPVSSEMNQLYTRADM 250
        GFDYLDRDNMVV +E+MVQRPL +A++DEVDS+L+DEARTPLI+SG + + LY RA+
    Sbjct: 181 GFDYLDRDNMVVYKEEMVQRPLAFAVIDEVDVSLVDEARTPLIISGE-AEKSTILYVRANT 239

35   Query: 251 FVKTL-NSDDYIIDVPTKTIGLSDTGIDKAENYFHLNLYDLENVALTHYIDNALRANYI 309
        FV+TL +DY +D+ TK++ L++ G+ K ENYF + NL+DLEN + H+I AL+ANY
    Sbjct: 240 FVRTLTBBDYTVDIKTKSVQLTEDGMTKGENYFDVENLFDLENTVILHHIAQALKANYT 299

    Query: 310 MLLNIDYVVSSEEQELIVDQFTGRTEGRRFSDGLHQAIEAKESVPIQEESSKTSASITYQ 369
        M L++DYVV ++ E+LIVDQFTGR M+GRRFS+GLHQA+EAKE V IQ ESKT A+IT+Q
40   Sbjct: 300 MSLDVDYVYV-QDDEVLIVDQFTGRIMKGRRFSEGLHQAIEAKEGVTIQNESKTMATITFQ 358

    Query: 370 NMFRMYHKLAMTGTGTGTEEEEFREIYNMRVPIPTNRPVQRIDHSDLLYPTLDSKFRAV 429
        N FRMY KLAGMTGT KTEEEEFRIYNMRVI IPTN+ + R D DL+Y T+++KF AV
45   Sbjct: 359 NYFRMYKLAGMTGTAKTEEEEFRIYNMRVIEIPTNKVIRDDRPDLIYTTMEAKFNAV 418

    Query: 430 VADVKEYEQGQPVVLVGTVAVETSDLISRKLVAAGVPHEVLNAKNHFKEAQIIMNAGQRG 489
        V D+ ER+ +GQPVVLVGTVA+ +LIS KL G+ H+VLNAK H +EA II +AG+RG
50   Sbjct: 419 VEDIAERHAKGQPVVLVGTVAMNI-ELISSKLKRKGIKHVDLNAKQHEREADIKHHAGERG 477

    Query: 490 AVTIATNMAGRGTDIKLGEGVRELGGLCVIGTERHESRRIDNQLRGRSGRQGDGPESQFY 549
        AV IATNMAGRGTDIKLGEG E GGL VIGTERHESRRIDNQLRGRSGRQGDGP +QFY
    Sbjct: 478 AVVIATNMAGRGTDIKLGEGTIEAGGLAVIGTERHESRRIDNQLRGRSGRQGDGPVTOFY 537

55   Query: 550 LSLEDDLMRRFGTDRIKVVLERMNLAEDDTVIKSKMLTRQVESAQRRVEGNNYDTRKQVL 609
        LS+ED+LMRRFG+D +K ++ER +AED I+SKM++R VESAQRRVEGNN+D+RKQVL
    Sbjct: 538 LSMEDELMMRRFGSDNMKSMMERFGMAED--AIQSKMVSRAVESAQRRVEGNNFDSRKQVL 595

    Query: 610 QYDDVMREQREIYYANRREVITAERDLGPELKGMIKRTIKRAVDAHSRSDKNTAA---EA 666
        QYDDV+R+QRE+IY R EVI AE L ++ MI+RT+ V +++ S + A +
60   Sbjct: 596 QYDDVLRQREVIYKQRYEVINAENSLREIIEQMIQRTVNFVSSNASSHEPEEAWNLLQ 655

```

-1305-

Query: 667 IVNFARSALLDEEAITVSELRLKEAEIKELLYERALAVYEQQIAKLKDPEATIEFQKVL 726
 I+++ + LL E IT+ +L+ +I+ L+ ++ A Y+++ L PE EF+KV+
 Sbjct: 656 IIDYVDANLLPEGTITLEDLQRTSEDIQNLI LDKIKAAYDEK-ETLLPPEEFNEFEKVV 714

5 Query: 727 ILMVVDNQWTEHIDALDQLRNSVGLRGYAQNNPIVEYQSEGFRMFQDMIGSIEFDVTRTL 786
 +L VVD +W +HIDA+D LR+ + LR Y Q +P+ EYQSEGF MF+ M+ SI+ DV R +
 Sbjct: 715 LLRVVDTKVVDHIDAMDHLRDGIHLRAYGQIDPLREYQSEGFMFEAMVSSIDEDVARYI 774

10 Query: 787 MKAQIHEQ-ERER-ASQHATTTAEQNISAQHVPMMNSESPEYQGIKRNDCPCGSGMKFKN 844
 MKA+I + ERE+ A A AE A+ P+ + Q I RND CPCGSG K+KN
 Sbjct: 775 MKAIRQNLEREQVAKGEAINPAEGKPEAKRQPIRKD----QHIGRNDPCPCGSGKKYKN 830

Query: 845 CHG 847
 CHG
 15 Sbjct: 831 CHG 833

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3619> which encodes the amino acid sequence <SEQ ID 3620>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4443(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 710/837 (84%), Positives = 777/837 (92%), Gaps = 3/837 (0%)

30 Query: 11 MANILRTVIENDKGELKKLDKIAKKVDSYADHMAALSDEALQAKTPEFKERYQNGETLDQ 70
 MANILR VIENDKGEL+KL+KIAKKV+SYAD MA+LSD LQ KT EFKERYQ GETL+Q
 Sbjct: 1 MANILRKVIENDKGELRKLEKIAKKVESYADQMASLSDRDLQGTLEFKERYQKGETLEQ 60

35 Query: 71 LLPEAFVAVVREASKRVLGLYPYHVQIMGGIVLHHGDIPEMRTGEGKTLTATMPVYLNAIS 130
 LLPEAFVAVVREA+KRVLGL+PY VQIMGGIVLH+GD+PEMRTGEGKTLTATMPVYLNAI+
 Sbjct: 61 LLPEAFVAVVREAAKRVLGLFPYRVQIMGGIVLHNGDVPPEMRTGEGKTLTATMPVYLNAIA 120

40 Query: 131 GLGVHVVITVNEYLSTRDATEMGEVYSWLGLSVGINLAAKSPFEKREAYNC DITYSTNAEV 190
 G GVHVVITVNEYLSTRDATEMGEVYSWLGLSVGINLAAKSP EKREAYNC DITYSTN+EV
 Sbjct: 121 GEGVHVVITVNEYLSTRDATEMGEVYSWLGLSVGINLAAKSPA EKREAYNC DITYSTNSEV 180

45 Query: 191 GFDYLRDNMVMVRQEDMVQRPLN+ALVDEVDVSLIDEARTPLIVSGPVSSEM NQLYTRADM 250
 GFDYLRDNMVMVRQEDMVQRPLN+ALVDEVDVSLIDEARTPLIVSG VSSE NQLY RADM
 Sbjct: 181 GFDYLRDNMVMVRQEDMVQRPLN+ALVDEVDVSLIDEARTPLIVSGAVSSETNQLYIRADM 240

50 Query: 251 FVKTLNSDDYIIDVPTKTIGLSDTGIDKAENYFHLNNLYDLENVALTHYIDNALRANYIM 310
 FVKTL S DY+IDVPTKTIGLS+GIDKAE+YF+L+NL+YD+ENVALTH+IDNALRANYIM
 Sbjct: 241 FVKTLTSVDYVIDVPTKTIGLSDSGIDKAESYFNLSNLYDIENVALTHFIDNALRANYIM 300

55 Query: 371 MFRMYHKLAMGTGTGKTEEEEFREIYNMRVPIPIPTNRPVQRIDHSDLLYPTLDSKFRAVV 430
 MFRMY KLAGMTGT KTEEEEFRE+YNMR+IPIPTNRP+ RIDH+DLLYPTL+SKFRAVV
 Sbjct: 361 MFRMYKKLAGMTGTAKTEEEEFREVYNMRIPIPIPTNRIARIDHTDLLYPTLESKFRAVV 420

60 Query: 431 ADVKERYEQGPVLVGTVAVETSDLISRKLVAAGVPHEVLNAKNHFKEAQIIMNAGQ RGA 490
 DVK R+ +GQP+LVGTVAVETSDLISRKL V AG+PHEVLNAKNHFKEAQIIMNAGQ RGA
 Sbjct: 421 EDVKTFRHAKGQPILVGTVAVETSDLISRKLVEAGIPHEVLNAKNHFKEAQIIMNAGQ RGA 480

65 Query: 491 VTIATNMAGRGTDIKLGEGVRELGGLCVIGTERHESRRIDNQLRGRSGRQGDPGESQFY L 550
 VTIATNMAGRGTDIKLGEGVRELGGLCVIGTERHESRRIDNQLRGRSGRQGDPGESQFY L
 Sbjct: 481 VTIATNMAGRGTDIKLGEGVRELGGLCVIGTERHESRRIDNQLRGRSGRQGDPGESQFY L 540

-1306-

Query: 551 SLEDDLRRFGTDRIKVVLERMNLAEEDTVIKSKMLTRQVESAQRRVEGNNYDTRKQVLQ 610
 SLEDDLRRFG+DRIK L+RM L E+DTVIKS ML RQVESAQ+RVEGNNYDTRKQVLQ
 Sbjct: 541 SLEDDLRRFGSDRIKAFDRMKLDEEDTVIKSGMLGRQVESAQKRVEGNNYDTRKQVLQ 600

5 Query: 611 YDDVMREQREIIYANRREVITAERDLGPGLKGMKRTIKRAVDAHRSRDKNTAAEAI VNF 670
 YDDVMREQREIIYANRR+VITA RDLGPE+K MIKRTI RAVDAH+RS++ A +AIV F
 Sbjct: 601 YDDVMREQREIIYANRRDVITANRDLGPEIKAMIKRTIDRAVDAHARSNRKDAIDAIVTF 660

10 Query: 671 ARSALLDEEAITVSELRLGLKEABIKELLYERALAVYEQQIAKLKDPEAIEFQKVLILMV 730
 AR++L+ EE I+ ELRGLK+ +IKE LY+RALA+Y+QQ++KL+D EAIIEFQKVLILM+
 Sbjct: 661 ARTSLVPEEFISAKELRGLKDDQIKEKLYQRALAIYDQQLSKLRDQEAIEFQKVLILMI 720

15 Query: 731 VDNQWTEHIDALDQLRNSVGLRGYAQNNPIVEYQSEGFRMFQDMIGSIEFDVTRTLMKAQ 790
 VDN+WTEHIDALDQLRN+VGLRGYAQNNP+VEYQ+EGF+MFQDMIG+IEFDVTRT+MKAQ
 Sbjct: 721 VDNKQWTEHIDALDQLRNAVGLRGYAQNNPVVEYQAEQGFQDMIGAEFDVTRTMMKAQ 780

Query: 791 IHEQERERASQHATTTAEQNISAQHVPMMNESPEYQGIKRNKPCGSGMKFKNCHG 847
 IHEQERERASQ ATT A QNI +Q ++ P+ ++RN+ CPCGSG KFKNCHG
 Sbjct: 781 IHEQERERASQRATTAAPQNIQSQQSANTDDLKPK---VERNEACPCGSGKKFKNCHG 834

20

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1166

A DNA sequence (GBSx1242) was identified in *S.agalactiae* <SEQ ID 3621> which encodes the amino acid sequence <SEQ ID 3622>. This protein is predicted to be phospho-2-dehydro-3-deoxyheptonate aldolase (aroH). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3429(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF40753 GB:AE002387 phospho-2-dehydro-3-deoxyheptonate
 aldolase, phe-sensitive [Neisseria meningitidis MC58]
 Identities = 122/348 (35%), Positives = 187/348 (53%), Gaps = 32/348 (9%)

40 Query: 1 MGFHQLSDKINIEILKQKTSLDLEVSQKKLAKL-----EELKNIKGEDQRFLVIV 51
 M H +D I I+ +K+ + + ++KE +E+ +++ G D+R LVI+
 Sbjct: 1 MTHHYPTDDIKIKEVKELLPPIAHLYELPISKEASGLVHRTRQEISDLVHGRDKRLVII 60

45 Query: 52 GPCADNPKAFLTYAKRLAKLEAFKDKMFLVMRVYTAKPRTNGDGYKGLVHSDKLGVF 111
 GPCS +PKA L YA+RL KL +++++ +VMRVY KPRT G+KGL++ G F
 Sbjct: 61 GPCSIHDPKAALEYAERLLKLRQYENELLIVMRVYFEKPRTT-VGWKGLINDPHLDGTF 119

50 Query: 112 -----FQARKMHYDIIRETGLLTADELLYPEMLSVMDLVSYYAIGARSVEDQGHFIS 165
 QAR + + G+ + E L DL+S+ AIGAR+ E Q HR ++
 Sbjct: 120 DINFGRLQARSLLS-LNNMGPASTFELDMITPQYYADLISWGAIGARTTESQVHRELA 178

55 Query: 166 SGIDAPVGMKNPTSGNLRVMFNAVYAAQNQOELFYQNKQ-----VRTDGNLLSHVILRGY 220
 SG+ PVG KN T GNL++ +A+ AA + K V T GN HVILRG
 Sbjct: 179 SGLSCPVGPKNGTDGNLKIADAIGAASHSHHFLSVTKAGHSAIVHTGGNPDCHVILRGG 238

Query: 221 HNADYRSIPNYHYENLLETITHYEETDLQNPFI VVDTNHDNSGKQFLEQIRIVKSVLADR 280
 PNY E++ E + + +++D +H NS K + Q+ + + + A
 Sbjct: 239 KE-----PNYDAEHVSEAAEQLRAGVTDK-LMIDCSHANSRKDYTRQMEVAQDIAAQL 291

60 Query: 281 QWHTKIRNYVRGFLIESYLEDGRODKPDVFGKSITDPCLGWDKTEMLI 328
 + + + G ++ES+L +GRQDKP+V+GKSITD C+GW TE L+

-1307-

Sbjct: 292 E---QDGGNIMGVMVESHLVEGRQDKPEVYVGKSITDACIGWGATEELL 336

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3623> which encodes the amino acid sequence <SEQ ID 3624>. Analysis of this protein sequence reveals the following:

5 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1171(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 52/233 (22%), Positives = 93/233 (39%), Gaps = 40/233 (17%)

15 Query: 50 IVGPCSADNPKAVLTYAKRLAKLEAAFKDKMFLVMRVYTAKPRTNGDGYKGLVHHSDKLG 109
 IVGPCS ++ + A KL + R KPRT+ ++GL
 Sbjct: 19 IVGPCSIESYDHIRLAASSAKKLGYN-----FRGGAYKPRTSAASFQGLG----- 64

20 Query: 110 VFFQARKMHYDIIRETGLLTADLLYPEMLSVMDLVSYYAIGARSVEDQGHRFISSGID 169
 Q + +++ +E GLL+ E++ L D + +GAR++++ S ID
 Sbjct: 65 --LQGIRYLHEVCQEFGLLSVSEIMSERQLEEAAYDYLVDVIQVGARNMQNFEFLKTLSHID 122

25 Query: 170 APVGMKNPTSGNLRVMFNAVYAAQNQQELFYQNKQVRTDGNLLSHVIL--RGYHNADYRS 227
 P+ K + A+ Q+ + S++IL RG D
 Sbjct: 123 KPILFKRGLMATIEEYLGALSYLQDTGK-----SNIILCERGVRGYD--- 164

30 Query: 228 IPNYHYENLLETITHYEETDLQNPFIIVVDTNHDNSGKQ-FLEQIRIVKSVLAD 279
 + + +++ ++TDL I+VD +H + L +I K+V A+
 Sbjct: 165 VETRNMLDIMAVPIIQKTDLP---IIVDVSHSTGRRDLLLPAAKIAKAVGAN 214

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1167

35 A DNA sequence (GBSx1243) was identified in *S.agalactiae* <SEQ ID 3625> which encodes the amino acid sequence <SEQ ID 3626>. This protein is predicted to be AcpS (acpS). Analysis of this protein sequence reveals the following:

40 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.3620(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG22706 GB:AF276617 acyl carrier protein synthase; AcpS
 [Streptococcus pneumoniae]
 50 Identities = 61/117 (52%), Positives = 90/117 (76%), Gaps = 1/117 (0%)

 Query: 1 MIVGHGIDLQIEIAITKAYERNQFAERVLTEQELLFPGKISNPKRQMSFLTGRWAAKEA 60
 MIVGHGID++E+ +I A R++ FA+RVLT QE+ F + +RQ+ +L GRW+AKEA
 Sbjct: 1 MIVGHGIDIBELASIESAVTRHEGFQAKRVLTAQEMERFTSLKG-RRQIEYLAGRWSAKEA 59

55 Query: 61 YSKALGTGIGKVNFDIEILSDDKGAPLITKEPFNGKSFVSISHSGNYAQASVILEE 117
 +SKA+GTGI K+ F D+E+L++++GAP ++ PF+GK ++SISH+ + ASVILEE
 Sbjct: 60 FSKAMGTGISKLGFDLEVLNNERGAPYFSQAPFSGKIWLSISHTDQFVTASVILEE 116

-1308-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3627> which encodes the amino acid sequence <SEQ ID 3628>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2001(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 76/119 (63%), Positives = 99/119 (82%), Gaps = 1/119 (0%)

Query: 1 MIVGHGIDLQEIEAITKAYERNQRFARVLTEQELLLFKGISNPKRQMSFLTGRWAAKEA 60

MIVGHGIDLQEI AI K Y+RN RFA+++L TEQEL +F+ KR++++L GRW+ KEA

Sbjct: 1 MIVGHGIDLQEISAIEKVYQRNPRFAQKILTEQELAFESFPY-KRRLNYLAGRWSGKEA 59

Query: 61 YSKALGTGIGKVNPHDIEILSDDKGAPLITKEPFNGKSFVSISHSGNYAQASVILEEEK 119

++KA+GTGIG++ F DIEIL+D +G P++TK PF G SF+SISHSGNY QASVILE++K

Sbjct: 60 FAKAIGTGIGRLTFQDIEILNDVRGCPILTKSPFKGNSFISISHSGNYVQASVILEDKK 118

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1168

A DNA sequence (GBSx1244) was identified in *S.agalactiae* <SEQ ID 3629> which encodes the amino acid sequence <SEQ ID 3630>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.24 Transmembrane 78 - 94 (77 - 97)

----- Final Results -----

bacterial membrane --- Certainty=0.2296(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD51027 GB:AF171873 alanine racemase [Streptococcus pneumoniae]

Identities = 227/366 (62%), Positives = 270/366 (73%)

Query: 1 MISSYHRPTRALIDLEAIANNVKSQEHIPSDKKTFAVVKANAYGHGAVEVSKYIESIVD 60

M+S HRPT+ALI L AI N++ + HIP AVVKANAYGHGAV V+K I+ VD

Sbjct: 1 MKASPHRPTKALIHGAIHQNIQQMGAIHPQGTLLKLA VVKANAYGHGAVAVAKAIQDDVD 60

Query: 61 GFCVSNLDEAIELRQAGIVKMILVLGVMPPEQVILAKNENITLTVASLEWLRLCQTSADV 120

GFCVSN+DEAIELRQAG+ K IL+LGV E V LAK + TLTVA LEW++ VD

Sbjct: 61 GFCVSNIDEAIELRQAGLSKPILILGVSEIEAVALAKEYDFTLTVAGLEWIQALLDKEVD 120

Query: 121 LSGLEVHIKVDSCMGRIQVRLDEGNKLISELGESGASVKGIFTHFATADEADNCKFNQ 180

L+GL VH+K+DSGMGRIG R+ E + L + G V+GIFTHFATADE + FN Q

Sbjct: 121 LTGLTVHLKIDSCMGRIQFREASEVEQAQDLLQOHGVCVEGIFTHFATADEESDDYFNAQ 180

Query: 181 LTFPKDFISGLDNCPLVHASNSATSLWHSETIFNAVRLGVVMYGLNPSGTDLDLPYPIN 240

L FK ++ + P+LVHASNSAT+LWH ETIFNAVR+G MYGLNPSG LDLPY +

Sbjct: 181 LERPKTILASMKEVPPELVHASNSATTLWHVETIFNAVRMGDAMYGLNPSGAVLDLPYDLI 240

Query: 241 PALSLESELVHVKQLHDGSGQVGYGATYQVTGDEFVGTVPVIGYADGWTRDMQGFVIVNGE 300

PAL+LES LVHVK + G+ +GYGATYQ ++ + TVPVGADGWTRDMQ FSV+V+G+

Sbjct: 241 PALTLESALVHVKTVPAGACMGYATYQADSEQVIATVPVIGYADGWTRDMQNFVSLVDGQ 300

-1309-

Query: 301 LCEIIGRVSMQMTIRLPQKYTIGTKVTLIGQQGSCNITTTDVAQKQRTINYEVLCLLSD 360
 C I+GRVSMQ+TIRLP+ Y +GTKVTLIG G IT T VA R TINYEV+CLLSD
 Sbjct: 301 ACPIVGRVSMQITIRLPKLYPLGTVTLIGSNGDKETATQVATYRVTTINYEVVCLLSD 360

5 Query: 361 RIPRY 366
 RIPR Y
 Sbjct: 361 RIPREY 366

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3631> which encodes the amino acid
 10 sequence <SEQ ID 3632>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.34 Transmembrane 82 - 98 (82 - 98)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the databases:

>GP:AAD51027 GB:AF171873 alanine racemase [Streptococcus pneumoniae]
 Identities = 222/366 (60%), Positives = 273/366 (73%)

25 Query: 1 MISSFHRPTVARVNLQAIKENVASVQKHIPLGVKTYAVVKADAYGHGAVQVSKALLPQVD 60
 M +S HRPT A ++L AI++N+ + HIP G AVVKA+AYGHGAV V+KA+ VD
 Sbjct: 1 MKASPHRPTKALIHLAGAIRQNIQQMGAIHPQGTLLKAVVKANAYGHGAVAVAKAIQDDVD 60

30 Query: 61 GYCVSNLDEALQLRQAGIDKEILILGVLLPNELELAVANAITVTIASLDWIALARLEKKE 120
 G+CVSN+DEA++LRQAG+ K ILILGV + LA T+T+A L+WI ++ +
 Sbjct: 61 GFCVSNIDEAIELRQAGLSKPIILILGVSEIEAVALAKEYDFTLTVAGLEWIQALLDKEVD 120

35 Query: 121 CQGLKVHVKVDSGMGRIGLRSSKEVNLLIDSLKELGADVEGIFTHFATADEADDTKFNQQ 180
 GL VH+K+DSGMGRIG R + EV D L++ G VEGIFTHFATADE D FN Q
 Sbjct: 121 LTGLTVHLKIDSGMGRIGFREASEVEQAQDLLQOHGVCVEGIFTHFATADEESDDYFNAQ 180

40 Query: 181 LQFFKKLIAGLEDKPRLVHASNSATSIIWHSDTIFNAVRLGIVSYGLNPSGSDLSLFFPLQ 240
 L+ FK ++A +++ P LVHASNSAT++WH +TIFNAVR+G YGLNPSG+ L LP+ L
 Sbjct: 181 LERFKTILASMKEVPELVHASNSATTLWHVETIFNAVRMGDAMYGLNPSGAVLDLPYDLI 240

45 Query: 241 EALSLESSLVHVKMISAGDTVGYGATYTAKKSEYVGTVPPIGYADGWTRNMQGFVSLVDGQ 300
 AL+LES+LVHVK + AG +GYGATY A + + TVPIGYADGWTR+MQ FSVSLVDGQ
 Sbjct: 241 PALTLESALVHVKTVPAGACMGYGATYQADSEQVIATVPPIGYADGWTRDMQNFSVSLVDGQ 300

50 Query: 301 FCEIIGRVSMQMTIRLPKAYPLGTVTLIGSNQKNISTTDIANRNTINYEVLCLLSD 360
 C I+GRVSMQ+TIRLPK YPLGTVTLIGSN K I+ T +A YR TINYEV+CLLSD
 Sbjct: 301 ACPIVGRVSMQITIRLPKLYPLGTVTLIGSNGDKETATQVATYRVTTINYEVVCLLSD 360

55 Query: 361 RIPRIY 366
 RIPR Y
 Sbjct: 361 RIPREY 366

An alignment of the GAS and GBS proteins is shown below.

Identities = 247/366 (67%), Positives = 295/366 (80%)

55 Query: 1 MISSYHRPTRALIDLEAIANNVKSQVEHIPSDKKTFAVVKANAYGHGAVEVSKYIESIVD 60
 MISS+HRPT A ++L+AI NV SVQ+HIP KT+AVVKA+AYGHGAV+VSK + VD
 Sbjct: 1 MISSFHRPTVARVNLQAIKENVASVQKHIPLGVKTYAVVKADAYGHGAVQVSKALLPQVD 60

60 Query: 61 GFCVSNLDEAIELRQAGIVKMILVLGVVMPEQVILAKNENITLTVASLEWLRCLQTSADV 120
 G+CVSNLDEA++LRQAGI K IL+LGV++P ++ LA IT+T+ASL+W+ L + +
 Sbjct: 61 GYCVSNLDEALQLRQAGIDKEILILGVLLPNELELAVANAITVTIASLDWIALARLEKKE 120

Query: 121 LSGLEVHIKVDSGMGRIGVRQLDEGNKLISELGESGASVKGIFTHFATADEADNCKFNQQ 180
 GL+VH+KVDSGMGRIG+R E N LI L E GA V+GIFTHFATADEAD+ KFNQQ

-1310-

Sbjct: 121 CQGLKVHVKVDSCMGRIGLRSSKEVNLLIDSLKELGADVEGIFTHFATADEADDTKFNQQ 180

Query: 181 LTFFKDFISGLDNCPLVHASNSATSLWHSETIFNAVRLGVVYGLNPSGTDLDLPYPIN 240
L FFK I+GL++ P LVHASNSATS+WHS+TIFNAVRLG+V YGLNPSG+DL LP+P+

5 Sbjct: 181 LQFFKKLIAGLEDKPRLLVHASNSATSIWHSDTIFNAVRLGIVSYGLNPSGSDLSLPPFLQ 240

Query: 241 PALSLESELVHVQLHDGSQVGYGATYQVTGDEFVGTVPVIGYADGWTRDMQGFSVIVNGE 300
ALSLES LVHVK + G VGYGATY E+VGTVPVIGYADGWTR+MQGFSV+V+G+

10 Sbjct: 241 EALSLESSLVHVKMISAGDTVGYGATYTAKKSEYVGTVPVIGYADGWTRNMQGFSVLVDGQ 300

Query: 301 LCEIIGRVSMQMTIRLPQKYTIGTKVTLIGQQGSCNITTTDVAQKRQTINYEVLCLLSD 360
CEIIGRVSMQ+TIRLP+ Y +GKVTIG NI+TTD+A R TINYEVLCLLSD

15 Sbjct: 301 FCEIIGRVSMQDLTIRLPKAYPLGKVTIGSNQKNISTTDIANYRNTINYEVLCLLSD 360

Query: 361 RIPRY 366
RIPR Y

Sbjct: 361 RIPRIY 366

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1169

A DNA sequence (GBSx1245) was identified in *S.agalactiae* <SEQ ID 3633> which encodes the amino acid sequence <SEQ ID 3634>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

30 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 1988.

A related GBS gene <SEQ ID 8745> and protein <SEQ ID 8746> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 4
McG: Discrim Score: 8.81
GvH: Signal Score (-7.5): 0.659999
Possible site: 27
40 >>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 1.06 threshold: 0.0
PERIPHERAL Likelihood = 1.06 247
modified ALOM score: -0.71

45 *** Reasoning Step: 3

----- Final Results -----

50 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear)

SEQ ID 8746 (GBS98) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 5; MW 80kDa).

GBS98-His was purified as shown in Figure 192, lane 9.

-1311-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1170

A DNA sequence (GBSx1246) was identified in *S.agalactiae* <SEQ ID 3635> which encodes the amino acid sequence <SEQ ID 3636>. This protein is predicted to be junction specific DNA helicase (mmsA) (recG). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.16    Transmembrane    530 - 546 ( 530 - 546)

----- Final Results -----
          bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA90280 GB:Z49988 MmsA [Streptococcus pneumoniae]
Identities = 483/671 (71%), Positives = 568/671 (83%)

Query: 1   MLLQSPISNLKGFPGPKSAEKFQKLDIYTVEDLLLYYPFRYEDFKSKSVFDLVDGEKAVIT 60
          M L P+ L G GPKSAEK+ KL I ++DILLY+PFRYEDFK+K V +L DGEKAV++
Sbjct: 1   MNLHQPLHLVLPVGPGPKSAEKYAKLGIENLQDILLYFPFRYEDFKTKQVLELEDGEKAVLS 60

Query: 61  GLVVTTPANVQYYGFKRNRLSFKLRQGEAVLNVSFFNQPYLADKIELGQEVAVFGKWDATK 120
          G VVTPA+VQYYGFKRNRL F L+QGE V V+FFNQPYLADKIELG +AVFGKD K
Sbjct: 61  GQVVTTPASVQYYGFKRNRLRFSKQGEVVFVAVNFFNQPYLADKIELGATLAVFGKWDRAK 120

Query: 121 SAITGMKVLAQVEDDMQPVYRVAQGISQSTLIKAISAFEISAHLELKENLPATLLEKYR 180
          +++TGMKVLAQVEDD+QPVYR+AQGISQ++L+K IK+AF+ L ++ENLP +LL+KY+
Sbjct: 121 ASLTGMKVLAQVEDDLQPVYRLAQGISQASLVKVIKTAFDQGLDLLEENLPQSLLDKYK 180

Query: 181 LMGRSQAACLAMHFPKDI TEYKQALRRIKFEELFYFQMNQLVLSKSENKSETNGLPILYSKH 240
          LM R QA AMHFPK+ EYKQALRRIK ELFYFQ LQ LKSEN+ + +GL + +S+
Sbjct: 181 LMSRCQAVRAMHFPKDLAEYKQALRRIKFAELFYFQMLQTLKSENRVQSGGLVLNWSQE 240

Query: 241 AMETKISSLPFILTNQKRSLEDEILSDMSSGAHMNRLLQGDVGSGKTVIAGLSMYAAYTA 300
          + +SLPF LT AQ++SL EIL+DM S HMNRLLQGDVGSGKTIV+AGL+M+AA TA
Sbjct: 241 KVTAVKASLPFALTQAQEKSLQEILTDMKSDHMHNRLLQGDVGSGKTIVVAGLAMFAAVTA 300

Query: 301 GFQSALMVPTEILAEQHYISLQELFPDLSTAILTSGMKA AVKRTVLA AAIANGSVD MIVGT 360
          G+Q+ALMVPTEILAEQH+ SLQ LFP+L +A+LT +KAA KR VL IA G D+I+GT
Sbjct: 301 GYQAALMVPTEILAEQH FESLQNLFPNLKLALLTGSLKAEKREVLETTIAKGEADLIIGT 360

Query: 361 HALIQDSVQYHKLGLVITDEQHRFGVKQRRIFREKGENPDVLMMTATPIPRTLAITAFGE 420
          HALIQD V+Y +LGL+I DEQHRFGV QRR I REKG+NPDLVLMMTATPIPRTLAITAFG+
Sbjct: 361 HALIQDGV EYARLGLIIIDEQHRFGVGQRRILREKGDNPDLVLMMTATPIPRTLAITAFGD 420

Query: 421 MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGE LQKDAQVYVISPLIEESEALDLKN 480
          MDVSIID++PAGRKPI+TRW+KHEQL VL W++GE+QK +Q YVISPLIEESEALDLKN
Sbjct: 421 MDVSIIDQMPAGRKPIVTRWIKHEQLPQVLTWLEGEIQKGSQAYVISPLIEESEALDLKN 480

Query: 481 AVALHAELSTYFEGIAKVALVHGRMKNDKDAIMQDFKDKKSHILVSTTVIEVG VNV PNA 540
          A+AL EL+T+F G A+VAL+HGRMK+DEKD IMQDFK++K+ ILVSTTVIEVG VNV PNA
Sbjct: 481 AIALSEELTTHFAGKAEVALLHGRMKSDKDKIMQDFKERTDILVSTTVIEVG VNV PNA 540

Query: 541 TIMIIMDADRFGLSQLHQLRGRVGRGYQSYAVLVANPKTDSGKKRMTIMTETTDGFVLA 600
          T+MIIMDADRFGLSQLHQLRGRVGRG QSYAVLVANPKTDSGK RM IMTETT+GFVLA
Sbjct: 541 TVMIIMDADRFGLSQLHQLRGRVGRGDKQSYAVLVANPKTDSGKDRMRIMTETTINGFVLA 600

Query: 601 ESDLKMRSGSEIFGTRQSGIPEFQVADIVEDYPILEEARVASDIVKDNWKNTEWALI 660
          E DLKMRSGSEIFGTRQSG+PEFQVADI+ED+PILEEAR+VAS I W+E+ EW +I
Sbjct: 601 EEDLKMRSGSEIFGTRQSGLPEFQVADIIEDFPIL EARKVASIYISSIEAWQEDPEWRMI 660
```

-1312-

Query: 661 LDNLRQHSDFD 671
 +L + D
 Sbjct: 661 ALHLEKKEHLD 671

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3637> which encodes the amino acid sequence <SEQ ID 3638>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 530 - 546 (530 - 546)

----- Final Results -----

bacterial membrane --- Certainty=0.1065(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 641/671 (95%), Positives = 655/671 (97%)

Query: 1 MLLQSPISNLKGFPGPKSAEFQKLDIYTVEDLLLYYPFRYEDFKSKSVFDLVDGEKAVIT 60
 M+L +P+SNLKGFGPKSAEFQKLDIYTVEDLLLYYPFRYEDFKSKSVFDLVDGEKAVIT
 Sbjct: 1 MILTAPMSNLKGFPGPKSAEFQKLDIYTVEDLLLYYPFRYEDFKSKSVFDLVDGEKAVIT 60

Query: 61 GLVVTTPANVQYYGFKRNRLSFKLRQGEAVLNVSFFNQPYLADKIELGQEVAVFGKWDATK 120
 GLVVTTPANVQYYGFKRNRLSFKLRQGEAVLNVSFFNQPYLADKIELGQEVAVFGKWDATK
 Sbjct: 61 GLVVTTPANVQYYGFKRNRLSFKLRQGEAVLNVSFFNQPYLADKIELGQEVAVFGKWDATK 120

Query: 121 SAITGMKVLAQVEDDMQPVYRVAQGISQSTLIKAISAFEISAHLELKENLPATLLEKYR 180
 SAITGMKVLAQVEDDMQPVYRVAQGISQSTLIKAISAFEI AHLELKENLPATLLEKYR
 Sbjct: 121 SAITGMKVLAQVEDDMQPVYRVAQGISQSTLIKAISAFEIDAHELELKENLPATLLEKYR 180

Query: 181 LMGRSQAACLAMHFPKDI TEYKQALRRIKFEELFYFQMNQLVLSKSENKSETNGLPILYSKH 240
 LMGRSQAACLAMHFPKDI TEYKQALRRIKFEELFYFQMNQLVLSK+ENKSETNGLPILYSK
 Sbjct: 181 LMGRSQAACLAMHFPKDI TEYKQALRRIKFEELFYFQMNQLVLSKAENKSETNGLPILYSKR 240

Query: 241 AMETKISSLPFILTNQKRSLEDILSDMSSGAHMNRLLQGDVGSGKTVIAGLSMYAAYTA 300
 AMETKISSLPFILTNQKRSLED+ILSDMSSGAHMNRLLQGDVGSGKTVIAGLSMYAAYTA
 Sbjct: 241 AMETKISSLPFILTNQKRSLEDILSDMSSGAHMNRLLQGDVGSGKTVIAGLSMYAAYTA 300

Query: 301 GFQSALMVPTEILAEQHYISLQELFPDLSIAILTSGMKAAVKRTVLAAIANGSVDMIVGT 360
 GFQSALMVPTEILAEQHYISLQELFPDLSIAILTSGMKAAVKRTVLAAIANGSVDMIVGT
 Sbjct: 301 GFQSALMVPTEILAEQHYISLQELFPDLSIAILTSGMKAAVKRTVLAAIANGSVDMIVGT 360

Query: 361 HALIQDSVQYHKLGLVITDEQHRFGVKQRRIFREKGENPDVLMMTATPIPRTLAITAFGE 420
 HALIQDSVQYHKLGLVITDEQHRFGVKQRRIFREKGENPDVLMMTATPIPRTLAITAFGE
 Sbjct: 361 HALIQDSVQYHKLGLVITDEQHRFGVKQRRIFREKGENPDVLMMTATPIPRTLAITAFGE 420

Query: 421 MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGELQKDAQVYVISPLIEESEALDLKN 480
 MDVSIIDELPAGRKPI+TRWVKHEQLGTVLEWVKGELQKDAQVYVISPLIEESEALDLKN
 Sbjct: 421 MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGELQKDAQVYVISPLIEESEALDLKN 480

Query: 481 AVALHAELSTYFEGIAKVALVHGRMKNDEKDAIMQDFKDKKSHILVSTTVIEGVNVPNA 540
 AVALHAELSTYFEGIAKVALVHGRMKNDEKDAIMQDFKDKKSHILVSTTVIEGVNVPNA
 Sbjct: 481 AVALHAELSTYFEGIAKVALVHGRMKNDEKDAIMQDFKDKKSHILVSTTVIEGVNVPNA 540

Query: 541 TIMIIMDADRFGLSQLHQLRGRVGRGYKQSYAVLVANPKTDSGKKRMTIMTETTDGFVLA 600
 TIMIIMDADRFGLSQLHQLRGRVGRGYKQSYAVLVANPKTDSGKKRMTIMTETTDGFVLA
 Sbjct: 541 TIMIIMDADRFGLSQLHQLRGRVGRGYKQSYAVLVANPKTDSGKKRMTIMTETTDGFVLA 600

Query: 601 ESDLKMRGSGEIFGTRQSGIPEFQVADIVEDYPILIEEARVASDIVKDNWKNTEWALI 660
 ESDLKMRGSGEIFGTRQSGIPEFQVADIVEDYPILIEEAR+V++ IV D NW +W L+
 Sbjct: 601 ESDLKMRGSGEIFGTRQSGIPEFQVADIVEDYPILIEEARVSAIVSDPNWIYEKQWQLV 660

Query: 661 LDNLRQHSDFD 671
 N+R+ +D

-1313-

Sbjct: 661 AQNIRKKEVYD 671

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1171

A DNA sequence (GBSx1247) was identified in *S.agalactiae* <SEQ ID 3639> which encodes the amino acid sequence <SEQ ID 3640>. This protein is predicted to be aryl-alcohol dehydrogenase (b1647). Analysis of this protein sequence reveals the following:

```

10 Possible site: 50
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1562 (Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
15                bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10105> which encodes amino acid sequence <SEQ ID 10106> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP:BAB07646 GB:AP001520 aryl-alcohol dehydrogenase [Bacillus halodurans]
    Identities = 173/300 (57%), Positives = 224/300 (74%)

    Query: 7 IGQTGIQATRIALGCMRMSDLKKGQAEVVGTTALDLGINFFDHADIYGGGLSELRFDAI 66
        +G + ++ +A+GCMR++ + K+AE V TAL+ G NFFDHADIYGGG E F DAI
25 Sbjct: 6 LGSSSLEVPVVAVGCMRINAISKKEAERFVQTALEQGANFFDHADIYGGGECEEIFADAI 65

    Query: 67 KHLNVNRDKMIIQSKCGIREGYDFSKEYILSSVDGILERLGTGYLDLFLILHRPDVLVEP 126
        + R+K+I+QSKCGIREG FDFSKEYIL SVDGIL+RL T+YLD L+LHRPD LVEP
30 Sbjct: 66 QMNEAVREKIIQSKCGIREGRFDFSKEYILQSVDGILQRLKTDYLDVLLHRPDALVEP 125

    Query: 127 EEVAEAF TKLRAEGKVKHFGVSNQNRFMELLQSYLDEPLAVNQLQLSPAHTPMFDAGLN 186
        EEVAEAF I + GKV+HFGVSNQN Q+ELL+ ++ +P+ NQLQLS + M +G+N
35 Sbjct: 126 EEVAEAFDLLESSGKVRHFGVSNQNPQMIELLKKFVRQPIVANQLQLSITNATMISSGIN 185

    Query: 187 VNMLNKASIEHDDGIVDYCRLKRVTTIQAWSPFQIDLSRGLFVNHPDYKELNETIAKLAKN 246
        VNM N+++I D ++DYCRL VTIQ WSPFQ G+F+ + + ELN+ I +LA+
40 Sbjct: 186 VNMENESAINRDGSLDYCRLHDVTIQWSPFQYGFEGVFLGNDLFPFLNKKIDELA EK 245

    Query: 247 YNVSSEAIIVIAWILRHPAKMQAIVGSMNPSRLKAIDKANDIALTRKEWYDIYRSAGNILP 306
        Y VS+ I IAW+LRHPA MQ ++G+MN RLK KA++I LTR+EWY+IYR+AGNILP
40 Sbjct: 246 YEVSTTTIAIAWLLRHPANMQPVIGTMNLKRLKDCCKASEIRLTREWEYIYRAAGNILP 305

```

There is also homology to SEQ ID 780.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1172

A DNA sequence (GBSx1248) was identified in *S.agalactiae* <SEQ ID 3641> which encodes the amino acid sequence <SEQ ID 3642>. This protein is predicted to be shikimate 5-dehydrogenase (aroE) (aroE). Analysis of this protein sequence reveals the following:

```

50 Possible site: 21
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----

```

-1314-

bacterial cytoplasm --- Certainty=0.0988(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC74762 GB:AE000264 putative oxidoreductase [Escherichia coli K12]
 Identities = 114/279 (40%), Positives = 171/279 (60%), Gaps = 3/279 (1%)

10 Query: 10 ITGLIANPARHSLSPMLMNTSFQEKNNMYAYLTFEVEEGKLTAVRGVVRALGIRGVNVSM 69
 L GL+A P RHSLSP M N + ++ + + Y+ FEV+ A+ G++AL +RG VSM
 Sbjct: 9 LIGLMAYPIRHSLSPMQNKALEKAGLPFTYMAFEVDNDSFPGAIEGLKALKMRGTGVSM 68

15 Query: 70 PFQKQSVIPLDDLSPOAKLVGAVNTIVNQGGTGRLVGHMTDGIGCFKALAAQGFSAKNKI 129
 P KQ +D+L+P AKLVGA+NTIVN G R G+ TDG G +A+ GF K K
 Sbjct: 69 PNKQLACEYVDELTPAAKLVGAINTIVNDDGYLR--GYNTIDGTGHIRAIKESGFDIKGKT 126

20 Query: 130 ITIAGIGGSGKAVAVQAAMEGVAEIRLFNRNSSNYDKVIDLSDKIKKQFQIKVVVDYLEN 189
 + + G GG+ A+ Q A+EG+ EI+LFNR +DK + + ++ + V V L +
 Sbjct: 127 MVLGAGAGASTAIGAQAIEGLKEIKLFNRRDEFFDKALAFQVRNENTDCVVTVTDLAD 186

25 Query: 190 KTAFKDAIRTSHFYIDATSLGMRPLDNYSLINDPEILTPNLVVVDLVYKPKETALLRFVR 249
 + AF +A+ ++ + T +GM+PL+N SL+ND +L P L+V + VY P T LL+ +
 Sbjct: 187 QQAFAEALASADILTNGTKVGMKPLENESLVNDISLLHPGLLVTECVYNPHMTKLLQQAQ 246

Query: 250 QNGVKHAYNGLGMLIYQGAFAFQILITNQEMPISSEVERVL 288
 Q G K +G GML++QGAE F L T ++ P+ V++V+
 Sbjct: 247 QAGCK-TIDGYGMLLWQGAEOFTLWTGKDFPLEYVKQVM 284

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3643> which encodes the amino acid
 sequence <SEQ ID 3644>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have an uncleavable N-term signal seq

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

40 >GP:AAC74762 GB:AE000264 putative oxidoreductase [Escherichia coli]
 Identities = 132/280 (47%), Positives = 186/280 (66%), Gaps = 3/280 (1%)

45 Query: 11 LVSLLATPIRHSLSPKMHNEAYAKGLDYAYLAFEVGTEQLADAVQGIRALGIRGSNVSM 70
 L+ L+A PIRHSLSP+M N+A K GL + Y+AFEV + A++G++AL +RG+ VSM
 Sbjct: 9 LIGLMAYPIRHSLSPMQNKALEKAGLPFTYMAFEVDNDSFPGAIEGLKALKMRGTGVSM 68

50 Query: 71 PNKEAILPLDDLSPAELVGAVNTIVNKGKHLVGHITDGIGALRALADEGVSVKNKI 130
 PNK+ +D+L+PAA+LVGA+NT+VN DG +L G+ TDG G +RA+ + G +K K
 Sbjct: 69 PNKQLACEYVDELTPAAKLVGAINTIVNDDG--YLRGYNTIDGTGHIRAIKESGFDIKGKT 126

55 Query: 131 ITLAGVGGAGAKAIAVQLAFDGAKEVRLFNQATRLSSVQKLVTKLNQLTRTKVTQLDIED 190
 + L G GGA AI Q A +G KE++LFNR+ ++N+ T VI+ DL D
 Sbjct: 127 MVLGAGAGASTAIGAQAIEGLKEIKLFNRRDEFFDKALAFQVRNENTDCVVTVTDLAD 186

60 Query: 191 QTAFKEAIRESHLFIDATSVGMKPLENLSLITDPELIRPDLVVFIDIVYSPAETKLLAFAR 250
 Q AF EA+ + + + T VGMKPLEN SL+ D L+ P L+V + VY+P TKLL A+
 Sbjct: 187 QQAFAEALASADILTNGTKVGMKPLENESLVNDISLLHPGLLVTECVYNPHMTKLLQQAQ 246

Query: 251 QHGAQKQVINGLGMVLYQGAFAFLITGQDMPVDAIKPLLG 290
 Q G K I+G GM+L+QGAE F L TG+D P++ +K ++G
 Sbjct: 247 QAGC-KTIDGYGMLLWQGAEOFTLWTGKDFPLEYVKQVMG 285

An alignment of the GAS and GBS proteins is shown below.

-1315-

Identities = 166/288 (57%), Positives = 221/288 (76%)

Query: 4 LNGETLLTGLIANPARHSLSPLMWNTSFQEKNNMYAYLTFFVEEGKLTAVRGVRLGIR 63
 L+G TLL L+A P RHSLSP M N ++ + ++YAYL FEV +L +AV+G+RALGIR
 5 Sbjct: 5 LSGHTLLVSLLATPIRHSLSPKMHNEAYAKLGLDYAYLAFEVGTQLADAVQGIRALGIR 64

Query: 64 GVNVSMPFKQSVIPLDDLSPQAKLVGAVNTIVNQGGRVLVGHMTDGIGCFKALAAQGF 123
 G NVSMP K++++PLDDLSP A+LVGAVNT+VN+ G G LVGH+TDGIG +ALA +G
 10 Sbjct: 65 GSNVSMPNKEAILPLDDLSPAELVGAVNTVVKDGGKHLVGHITDGIGALRALADEGV 124

Query: 124 SAKNKIITTAGIGSGKAVAVQAAMEGVAEIRLFNRNSSNYDKVIDLSDKIKKQFQIKVV 183
 S KNKIIT+AG+GG+GKA+AVQ A +G E+RLFNRR ++ V L K+ + + KV
 15 Sbjct: 125 SVKNKIITLAGVGGAGKAIQVLAFAFDGAKVRLFNRRQATRLSSVQKLVTKLNQLTRTKVT 184

Query: 184 VDYLENKTAFAKDAIRTSHFYIDATSLGMRPLDNYSLINDPEILTPNLVVVDLVYKPKETA 243
 + LE++TAFK+AIR SH +IDATS+GM+PL+N SLI DPE++ P+LVV D+VY P ET
 20 Sbjct: 185 LQDLEDQTAFAKEAIRESHLFIDATSVGMKPLENLSLITDPELIRPDVVFIVYSPAETK 244

Query: 244 LLRFVRQNGVKHAYNGLGMLIYQGAEAFQLITNQEMPISSEVERVLQTE 291
 LL F RQ+G + NGLGM++YQGAEAF+LIT Q+MP+ +++ +L E
 25 Sbjct: 245 LLAFARQHGAQKVINGLGMVLYQGAEAFKLITGQDMPVDIAIKPLLGDE 292

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1173

A DNA sequence (GBSx1249) was identified in *S.agalactiae* <SEQ ID 3645> which encodes the amino acid sequence <SEQ ID 3646>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 30 INTEGRAL Likelihood = -6.16 Transmembrane 57 - 73 (53 - 76)

----- Final Results -----
 bacterial membrane --- Certainty=0.3463(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1174

A DNA sequence (GBSx1250) was identified in *S.agalactiae* <SEQ ID 3647> which encodes the amino acid sequence <SEQ ID 3648>. Analysis of this protein sequence reveals the following:

Possible site: 17
 45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2333(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10103> which encodes amino acid sequence <SEQ ID 10104> was also identified.

-1316-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05343 GB:AP001512 L-asparaginase [Bacillus halodurans]
Identities = 158/319 (49%), Positives = 214/319 (66%), Gaps = 4/319 (1%)

5  Query: 1  MKKILVLHTGGTISMNANEKGQVMSSADNPMKYVDLSLDDL-DLTVVDFLNLPSQITPH 59
      MKK+LV+HTGGTI+M+ +EKG V   NP+   SL  +  + V DFLN+PSP +TP
Sbjct: 1  MKKVLVIHTGGTIAMHEDEKGGVQPKETNPLFATVESLTSIASIEVDDFLNIPSPHMTPE 60

      Query: 60  HMLDIYHYLKQHASN--FDGVVITHGTDLTLEETAYFLDTMILPKIPIIITGAMRSTNELG 117
      M  +   LK   N   FDGVVITHGTDLTLEETAY LD ++  ++P+++TGAMRS+NELG
10  Sbjct: 61  LMFQLAERLKSRVGNESFDGVVITHGTDLTLEETAYLLDLLLDWEVPPVVTGAMRSSNELG 120

      Query: 118  SDGVYNYLSALRVANSTKAADKGVLVVMNDEIHAAYVTKTHTTNVSTFQTPHGPLGII 177
      +DG +N++SA++ A + +A  KGVLVV NDEIH AK VTKTHT+NV+TFQ+P +GP+GI+
15  Sbjct: 121  ADGPHNFISAVKTAATDEAKGKGVLVVFNDEIHTAKNVTKTHTSNVATFQSPQYGPPIGIV 180

      Query: 178  MKQDLLFFKATEERVFRFDLKITGTPIVKAYAGMGDSGIISFLNSQNISGLVIEALGAG 237
      K+ + F  A  +  +  +  I  V ++KAYAGM D  ++  +  I GLVIEA G G
20  Sbjct: 181  TKRGVTFHHAPSYESYTVSSIDHRVVLKAYAGM-DGSVVDIAIDTGDGLVIEAFQGG 239

      Query: 238  NMPPKAAQEIEELIEQGVVVLVSRFCNGIAEPVYGYEGGGAKLQESGVMFVKELNAPKA 297
      N+PP   I+ L +  +PVVLVSR +GI +  Y YEGGG L++ GV+F  LN KA
25  Sbjct: 240  NLPPAVVPSIKRLHQANIPVVLVSRSVSGIVQETYAYEGGGRHLKDLGVIFTNGLNGQKA 299

      Query: 298  RLKLLIALNAGLTGQNLKD 316
      RLKLL+AL   + L++
Sbjct: 300  RLKLLVALELTDRKKLQE 318
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3649> which encodes the amino acid sequence <SEQ ID 3650>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -2.28    Transmembrane    245 - 261 ( 243 - 261)

35  ----- Final Results -----
      bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB05343 GB:AP001512 L-asparaginase [Bacillus halodurans]
Identities = 158/320 (49%), Positives = 218/320 (67%), Gaps = 5/320 (1%)

45  Query: 1  MKKILVLHTGGTISMQADNSGRVVPNQDNPM-TKIHAAAQDIQLTVSDFLNLPSPHITPH 59
      MKK+LV+HTGGTI+M D G V P + NP+  +  +  + V DFLN+PSPH+TP
Sbjct: 1  MKKVLVIHTGGTIAMHEDEKGGVQPKETNPLFATVESLTSIASIEVDDFLNIPSPHMTPE 60

      Query: 60  HMLSIYHHIQERT--DVFDGIVITHGTDLTLEETAYFLDTMALPTNIPVVLTGAMRSSNEV 117
      M  +  ++ R  +  FDG+VITHGTDLTLEETAY LD + L  +PVV+TGAMRSSNE+
50  Sbjct: 61  LMFQLAERLKSRVGNESFDGVVITHGTDLTLEETAYLLDLL-LDWEVPPVVTGAMRSSNEL 119

      Query: 118  GSDGIYNYLTALRVASSDKAKEKGVLVVMNDEIHAAYVTKTHTTNISTFQTPHGPLGI 177
      G+DG +N++++A++ A++D+AK  KGVLVV NDEIH AK VTKTHT+N++TFQ+P +GP+GI
55  Sbjct: 120  GADGPHNFISAVKTAATDEAKGKGVLVVFNDEIHTAKNVTKTHTSNVATFQSPQYGPPIGI 179

      Query: 178  IMKNDLLFFKTAEPRIREFDLRCISGTIPIIKAYAGMGDSILSLTPGSIQGLVIEALGA 237
      + K  + F      +  +  +  I  + ++KAYAGM DGS++  +  I GLVIEA G
Sbjct: 180  VTKRGVTFHHAPSYESYTVSSIDHRVVLKAYAGM-DGSVVDIAIDTGDGLVIEAFQGG 238

      Query: 238  GNVPPPLAVGEIEHLIALGIPVILVSRFCNGMAEPVYAYEGGGAMLQEGVMFVKELNAPK 297
      GN+PP V I+ L  IPV+LVSR +G+ +  YAYEGGG L++ GV+F  LN K
60  Sbjct: 239  GNLPPAVVPSIKRLHQANIPVVLVSRSVSGIVQETYAYEGGGRHLKDLGVIFTNGLNGQK 298

      Query: 298  ARLKLLIALNAGLTGQELKD 317
```

-1317-

ARLKLL+AL ++L++
 Sbjct: 299 ARLKLLVALELT'TDRKKLQE 318

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 242/321 (75%), Positives = 275/321 (85%), Gaps = 1/321 (0%)

Query: 1 MKKILVLHTGGTISMNANEKGQVMSSADNPMKYVDLSLDDLDLTVVDFLNLPSQITPHH 60
 MKKILVLHTGGTISM A+ G+V+ + DNPM + + D+ LTV DFLNLPS ITPHH
 Sbjct: 1 MKKILVLHTGGTISMQADNSGRVVPNQDNPMTKIHAAQDIQLTVSDFLNLPSPHITPHH 60

10 Query: 61 MLDIYHYLKQHASNFDGVTHTGTDLEETAYFLDTMILP-KIPIIITGAMRSTNELGSD 119
 ML IYH++++ FDG+VITHGTDLEETAYFLDTM LP IP+++TGAMRS+NE+GSD
 Sbjct: 61 MLSIYHHIQERTDVFDGIVITHGTDLEETAYFLDTMALPTNIPVVLTGAMRSSNEVGSD 120

15 Query: 120 GVYNYLSALRVANSTKAADKGVLVVMNDEIHAAYVTKTHTTNVSTFQTPTHGPLGIIMK 179
 G+YNYL+ALRVA+S KA +KGVLVVMNDEIHAAYVTKTHTTN+STFQTPTHGPLGIIMK
 Sbjct: 121 GIYNYLTALRVASSDKAKEKGVLVVMNDEIHAAYVTKTHTTNISTFQTPTHGPLGIIMK 180

20 Query: 180 QDLLFFKATEERVRFDLDKITGTVPVIVKAYAGMGDSGIISFLNSQISGLVIEALGAGNM 239
 DLLFFK E R+REFDL I+GT+PI+KAYAGMGD I+S L +I GLVIEALGAGN+
 Sbjct: 181 NDLLFFKTAEPRIREFDLRCISGTIPIKAYAGMGDSILSLLTPGSIQGLVIEALGAGNV 240

25 Query: 240 PPKAAQEIEELIEQGVVVLVSRCFNGIAEPVYGYEGGAKLQESGVMFVKELNAPKARL 299
 PP A EIE LI G+PV+LVSRFCFNG+AEFVY YEGGGA LQE+GVMFVKELNAPKARL
 Sbjct: 241 PPLAVGEIEHLIALGIPVILVSRCFNGMAEPVYAYEGGAMLQEAGVMFVKELNAPKARL 300

30 Query: 300 KLLIALNAGLTGQNLKDYIEG 320
 KLLIALNAGLTGQ LKDYIEG
 Sbjct: 301 KLLIALNAGLTGQELKDYIEG 321

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1175

35 A DNA sequence (GBSx1251) was identified in *S.agalactiae* <SEQ ID 3651> which encodes the amino acid sequence <SEQ ID 3652>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4427(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAB85142 GB:AL162757 conserved hypothetical protein [Neisseria meningitidis Z2491]
 Identities = 87/285 (30%), Positives = 138/285 (47%), Gaps = 35/285 (12%)

50 Query: 4 KAVFFDIDGTLNDRKNVQKSTIK-AIRNLKDQGLVGLATGRG----PSFVQPFLENLG 58
 K VFFDID TL + + ++K A+ L+ +GIL LATGR P V+ + G
 Sbjct: 11 KIVFFDIDDTLYRKYTDTLRPSVKTAVAAALRGKGILTALATGRSLATPEKVRDMMMAETG 70

55 Query: 59 LDFAVTYNGQYIYSRSEIITYNQLSKTTVYRLIRYAGARRREISLGTASGLLGSGIIGLG 118
 +D VT NGQ+ + + + + R+ + SLG +G G+
 Sbjct: 71 MDAVVTINGQFALLHGKTVCEVPM DAGLMGRVCAHLD-----SLGMDYAFVVGGE--GIA 122

60 Query: 119 TSRLGQIVSSLVPRKWAKAIERSFKHFIRRIKPNIDSLMVILREPIYQVVLVATEGE-- 176
 S L + V R+ KH I +P+YQ+++ A E E
 Sbjct: 123 VSALSECVC-----RALKH----IASDFFADKDYFSSKPVYQMLVFAEENEMP 166

Query: 177 --SERIQKQFPRVKLTRSSPYSMVDVISEGQSKVKGIERVGQRYGFDLSEVIAFGDSNDI 234

-1318-

S+ +++++ +K R ++D++ G SK GI V + G +++++V+AFGD ND+
 Sbjct: 167 LWSDIVERE--GLKTVRWHEEAVDLLPAGASKTDGIRSVVEALGLEMAVMAFGDGLNDV 224

Query: 235 EMLSQVGIGVAMGNASQQVRENARYTTADNNDGISKALAHYGLI 279
 EMLS+VG GVAMGN Q +E A+Y ++DG+ + L G+I
 Sbjct: 225 EMLSEVGFGVAMGNGEQAAKEAAKYVCPGVDEEDGVLRGLQDLGVI 269

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3653> which encodes the amino acid sequence <SEQ ID 3654>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.6014(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 320/459 (69%), Positives = 391/459 (84%)

Query: 1 MAIKAVFFDIDGTLNDRKNVQKSTIKAIRNLKDQGILVGLATGRGPSFVQPFLENLGLD 60
 + +KAVFFDIDGTLNDRKN+QK+T KAI+ LK QGI+VGLATGRGP FVQPFLEN GLD
 Sbjct: 1 LTVKAVFFDIDGTLNDRKNIQKTTQKAIQQLKKQGIMVGLATGRGPGFVQPFLENFGLD 60

Query: 61 FAVTYNGQYIYSRSEIIYTNQLSKTTVYRLIRYAGARRREISLGTASGLLGSGIIGLGS 120
 FAVTYNGQYI +R +++Y NQL K+ +Y++IRYA ++REISLGTASGL GS II +GTS
 Sbjct: 61 FAVTYNGQYILTRDKVLYQNQLPKSMIYKVIRYANEKKREISLGTASGLAGSRIIDMGTS 120

Query: 121 RLQGVVSSLVPRKWKAKIERSFKHFIRRIKPQNIDSLMVILREPIYQVVLVATEGESERI 180
 GQ++SS VP+ WA+ +E SFKH IRRIKPQ+ +L+ I+REPIYQVVLVA++ E+++I
 Sbjct: 121 PFGQVISSFVPKSWARTVEGSEFKHLIRRIKPQSFRNLVTIMREPIYQVVLVASQAETKKI 180

Query: 181 QKQFPRVKLTRSSPYSMDVISEGQSKVKGIERVGQRYGFDLSEVIAFGDSNDNDIEMLSQV 240
 Q++FP +K+TRSSPYS+D+IS QSK+KGIER+G+ +GFDLSEV+AFGDSNDND+EMLS V
 Sbjct: 181 QEKFPPIKITRSSPYSLDLISVDQSKIKGIERLGEMFGFDLSEVMAFGDSNDNDLEMLSGV 240

Query: 241 GIGVAMGNASQQVRENARYTTADNNDGISKALAHYGLIQFEIEKTFSSRDENFNKVKSF 300
 GIG+AMGNA V++ A +TT NN+DGISKALAHYGLI F+IEK+F SRDENFNKVK F
 Sbjct: 241 GIGIAMGNAETVVKDGAHFTTDSNNNDGISKALAHYGLIHFDIEKSFKSRDENFNKVKDF 300

Query: 301 HLLMDGETIETPRLYDSKEAGFRSDFKVEEIVEFLYAASQGNQKVFQDSIRNLHLAIDKA 360
 H LMD +TIETPR Y EAG+RS FKVEEIVEFLYAAS+G+Q+ F Q+I +LH A+D+A
 Sbjct: 301 HRLMDSDTIETPRSYTISEAGYRSGFKVEEIVEFLYAASKGDQQQFTQAIQFDLHGAVDQA 360

Query: 361 RDKVISKDHDPETPLVGEVDALTDLLYLYTGSFVLMGVDPKPLFDTVHEANMGKIFPDGKA 420
 +KV +K H ETPL+G+VDAL DLY TYGSFVLMGVDP+P+F+ VHEANM KIFPDGKA
 Sbjct: 361 ANKVQAKKHVETPLIGQVDALADLLYFTYGSFVLMGVDPQPIFEAVHEANMAKIFPDGKA 420

Query: 421 HFDPVTHKILKPDWEEHFAPEPSIRRELDLSQIQKSLNR 459
 HFDPVTHKI KPD W+E APE +I++ELD Q+QKSL R
 Sbjct: 421 HFDPVTHKIQKPDYWQERHAPEVAIKKELDKQLQKSLQR 459

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1176

A DNA sequence (GBSx1252) was identified in *S.galactiae* <SEQ ID 3655> which encodes the amino acid sequence <SEQ ID 3656>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

-1319-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1671(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10101> which encodes amino acid sequence <SEQ ID 10102> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06903 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 61/141 (43%), Positives = 92/141 (64%)

Query: 22 YERILVAIDGSTESSELAFAEKAVNVALRNDSELILTHVIDTRALQSFATFDITYIEKLEKE 81
 Y ILVA+DGST+++ A KA N A ++L + HVID+R+ + +D + E +
 Sbjct: 2 YNHILVAVDGSTQAKRALYKAFNYAKEFKADLFICHVIDSRSFATVEQYDRTVVGAELD 61

Query: 82 AKDVLEEYEQAREKGADKVRQVIEFGNPKTLLAHDIPKEKVDLIMVGATGLNTFERFX 141
 K +L+ Y ++A + G DKV +++FG+PK ++ I +K +DLI+ GATGLN ERF
 Sbjct: 62 GKLLQRYSEEAEGAGVDKVHTILDFGSPKANISKTTIAQYDIDLIITGATGLNAVERFL 121

Query: 142 IGSSSEYILRHAKVDLLIVRD 162
 +GS SE + RHAK D+LIVR+
 Sbjct: 122 MGSVSESVARHAKCDVLIVRN 142

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3657> which encodes the amino acid sequence <SEQ ID 3658>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1296(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/156 (75%), Positives = 135/156 (86%)

Query: 12 LEEDRLMSQKYERILVAIDGSTESSELAFAEKAVNVALRNDSELILTHVIDTRALQSFATFD 71
 L+ED MS KY+RILVAIDGS ESELAF K VNALRND+ L+L HVIDTRALQS ATFD
 Sbjct: 25 LKEDSSMSLKYKRILVAIDGSYESELAFAKGVNVALRNDATLLLVHVIDTRALQSVATFD 84

Query: 72 TYIEKLEKEAKDVLEEYEQAREKGADKVRQVIEFGNPKTLLAHDIPKEKVDLIMVGA 131
 TYIEKLE+EAKDVL+++EKQA+ G ++Q+IEFGNPK LLAHDIP++E DLIMVGA
 Sbjct: 85 TYIEKLEQEAKDVLDDFEKQAQIAGITNIKQIEFGNPKNLLAHDIPDRENADLIMVGA 144

Query: 132 TGLNTFERFXIGSSSEYILRHAKVDLLIVRDPNKT 167
 TGLNTFER IGSSSEYI+RHAK+DLL+VRD KT+
 Sbjct: 145 TGLNTFERLLIGSSSEYIMRHAKIDLLVVRDSTKTL 180

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1177

A DNA sequence (GBSx1253) was identified in *S.agalactiae* <SEQ ID 3659> which encodes the amino acid sequence <SEQ ID 3660>. This protein is predicted to be aspartate aminotransferase (aspC). Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

-1320-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2803(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC21948 GB:U32714 aminotransferase [Haemophilus influenzae Rd]

Identities = 142/212 (66%), Positives = 181/212 (84%), Gaps = 1/212 (0%)

10

Query: 1 MKIFDKSMKLEHVAYDIRGPFVLEADRMRANGEKILRLNTGNPAAFGFEAPDEVIRDLIT 60

M++F KS KLEHV YDIRGPV +EA R+ G KIL+LN GNPA FGFEAPDE++ D++

Sbjct: 1 MRLFPKSDKLEHVCYDIRGPFVHKEALRLEEGNKILKLNIGNPAPFGFEAPDEILVDVLR 60

15

Query: 61 NARESEGYSDSKGIFSARKAVMQYYQLQNI-HVDMDDIYIVNGVSEGISMSMQALLDND 119

N ++GY DSKG++SARKA++QYYQ + I ++D+YI NGVSE I+M+MQALL++ D

Sbjct: 61 NLPSAQGYCDSKGLYSARKAIVQYQSKGILGATVNDVYIGNGVSELITMAMQALLNDGD 120

20

Query: 120 EVLVPMPPDYPLWTACVSLAGGNAVHYICDEEANWYPDIDDIKSKITSKTKAIVLINPNNP 179

EVLVPMPPDYPLWTA V+L+GG AVHY+CDE+ANW+P IDDIK+K+ +KTKAIV+INPNNP

Sbjct: 121 EVLVPMPPDYPLWTAAVTLGGKAVHYLCDEANWFPITIDDIKAKVNAKTKAIVLINPNNP 180

Query: 180 TGAVYPREILQEIVDIARQNDLIIFSDEVYDR 211

TGAVY +E+LQEIV+IARQN+LIIF+DE+YD+

25

Sbjct: 181 TGAVYSKELLQEIVEIARQNNLIIFADEIYDK 212

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3661> which encodes the amino acid sequence <SEQ ID 3662>. Analysis of this protein sequence reveals the following:

Possible site: 59

30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35

bacterial cytoplasm --- Certainty=0.2936(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/212 (80%), Positives = 193/212 (90%), Gaps = 1/212 (0%)

40

Query: 1 MKIFDKSMKLEHVAYDIRGPFVLEADRMRANGEKILRLNTGNPAAFGFEAPDEVIRDLIT 60

MKI +KS KLEHVAYDIRGPVL+EA+RM A+GEKILRLNTGNPAAFGFEAPDEVIRDLI

Sbjct: 13 MKIIEKSSKLEHVAYDIRGPFVLEADRMIASGEKILRLNTGNPAAFGFEAPDEVIRDLIV 72

45

Query: 61 NARESEGYSDSKGIFSARKAVMQYYQLQNI-HVDMDDIYIVNGVSEGISMSMQALLDND 119

NAR SEGYSDSKGIFSARKA+MQY QL+ VD++DIY+ NGVSE IS+S+QALLDN D

Sbjct: 73 NARLSEGYSDSKGIFSARKAIMQYQCQLKGFDPVDIEDIYLGNGVSELISISLQALLDNGD 132

50

Query: 120 EVLVPMPPDYPLWTACVSLAGGNAVHYICDEEANWYPDIDDIKSKITSKTKAIVLINPNNP 179

EVLVPMPPDYPLWTACVSL GG AVHY+CDEEA WYPDI DIKSKITS+TKAIV+INPNNP

Sbjct: 133 EVLVPMPPDYPLWTACVSLGGGKAVHYLCDEEAGWYPDIADIKSKITSRTKAIVVINPNNP 192

Query: 180 TGAVYPREILQEIVDIARQNDLIIFSDEVYDR 211

TGA+YP+EIL++IV +AR++ LIIF+DE+YDR

55

Sbjct: 193 TGALYPKEILEDIVALEHQLIIFADEIYDR 224

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1178

60 A DNA sequence (GBSx1254) was identified in *S.agalactiae* <SEQ ID 3663> which encodes the amino acid sequence <SEQ ID 3664>. Analysis of this protein sequence reveals the following:

-1321-

Possible site: 60

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -14.75 Transmembrane 38 - 54 (29 - 60)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.6901(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 A related GBS nucleic acid sequence <SEQ ID 9389> which encodes amino acid sequence <SEQ ID 9390> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3665> which encodes the amino acid sequence <SEQ ID 3666>. Analysis of this protein sequence reveals the following:

15 Possible site: 43

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -15.97 Transmembrane 35 - 51 (25 - 58)

20 ----- Final Results -----

bacterial membrane --- Certainty=0.7389(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 51/87 (58%), Positives = 63/87 (71%), Gaps = 7/87 (8%)

Query: 1 MAKKPWEKKVVENNSHRKDKITRTSRGVVSTPWITAFLSAFFVIVVAILFIVFYTSNRG 60
MAK+PWE+K+V++ + TR SR STPW+TA LS FFVI+VAILFI FYTSN G

30 Sbjct: 1 MAKEPWEKKIVDDTIGTR---TRKSRNAFISTPWLTALLSVFFVIIIVAILFIFFYTSNSG 57

Query: 61 EDRAKETSGFYGASSQKVNSSKTKKAS 87

+R ET+GFYGAS+ K KT+KAS

Sbjct: 58 SNRQAEITNGFYGASTHK----KTRKAS 80

35

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1179

40 A DNA sequence (GBSx1255) was identified in *S.agalactiae* <SEQ ID 3667> which encodes the amino acid sequence <SEQ ID 3668>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0815(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3669> which encodes the amino acid sequence <SEQ ID 3670>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

-1322-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0107 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 43/64 (67%), Positives = 53/64 (82%)

Query: 1 MKVALIPEKCIACGLCQTYSNIFDYQDDGIVKFSDDNLEKEIPSSDQDTVLAVKSCPTK 60
 MKV++IPEKCIACGLCQTY++FDY D+GIV FS + + I SD+D +LAVKSCPTK
 Sbjct: 1 MKVSIPEKCIACGLCQTYSSLFDYHDNGIVTFSSSSETSQSICPSDKDAILAVKSCPTK 60

Query: 61 ALTI 64

ALT+

Sbjct: 61 ALTL 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1180

A DNA sequence (GBSx1256) was identified in *S.agalactiae* <SEQ ID 3671> which encodes the amino acid sequence <SEQ ID 3672>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -10.61 Transmembrane 47 - 63 (41 - 69)

----- Final Results -----

bacterial membrane --- Certainty=0.5246 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC36851 GB:L23802 pore-forming peptide [Enterococcus faecalis]

Identities = 42/130 (32%), Positives = 63/130 (48%), Gaps = 9/130 (6%)

Query: 7 KIRYHWQPELSWAIYWSIAIAPIFIGLSLLYERTE---IPSQVFLFAIFIVLVGIGFH 63
 K +++WQPEL+ IYWS +FI L L E I + V V F +F L G
 Sbjct: 3 KQKFYQWPELASTIYWSCTFCILFISLILALENNGPYLISNLVMVFFVFAYL---GIA 59

Query: 64 RYFVIEEDGYLRIVSFNFLRRTKFPIEDIAKIEVTKSSVTIKFNNNHE--RIFYMRKWPK 121
 R F + E L + + R+ P+ I K+ + S+ I + E ++F M+K
 Sbjct: 60 RSFNMTETS-LIVRDVLWFRKKALPLSQIEKVTYNEKSIEIFSSEFKEGSKVFLMKKTD 118

Query: 122 KYFLDALAIE 131

FL+AL I+

Sbjct: 119 SLFLEALKIK 128

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3673> which encodes the amino acid sequence <SEQ ID 3674>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.87 Transmembrane 47 - 63 (41 - 69)

INTEGRAL Likelihood = -3.35 Transmembrane 20 - 36 (18 - 37)

----- Final Results -----

bacterial membrane --- Certainty=0.4949 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1323-

The protein has homology with the following sequences in the databases:

>GP:AAC36851 GB:L23802 pore-forming peptide [Enterococcus faecalis]
Identities = 42/130 (32%), Positives = 70/130 (53%), Gaps = 12/130 (9%)

5 Query: 7 KIRYHWQPELSWSIIYWSIAFAPIFVGLSLLYERTE---IPSRVFILFAIFAVLVGIGLH 63
K +++WQPEL+ +IIYWS F +F+ L L E I + V + F +FA L G+
Sbjct: 3 KQKFYWQPELASTIIYWSCTFCILFISLILALENNGPYLISNLVMVPFFVFAYL---GIA 59

10 Query: 64 RYF-IIENNGILRIVSFKLFGRKLLISTITKIEVTKSTLCL---HVEDKSYLFYMRKWP 119
R F + E + I+R V + F + L +S I K+ + + + ++ S +F M+K
Sbjct: 60 RSNMTTETSLIVRDVLW--FRKKALPLSQIRKVTYNEKSIEIFSSEFKEGSKVFLMKKKT 117

Query: 120 KKYFLDALAV 129
FL+AL +

15 Sbjct: 118 DSLFLEALKI 127

An alignment of the GAS and GBS proteins is shown below.

Identities = 115/162 (70%), Positives = 132/162 (80%), Gaps = 1/162 (0%)

20 Query: 1 MIKLFQKIRYHWQPELSWAIYWSIAIAPIFIGLSLLYERTEIPSQVFLFAIFIVLVGI 60
MIKLFQKIRYHWQPELSW+IIYWSIA APIF+GLSLLYERTEIPS+VF+LFAIF VLVGI
Sbjct: 1 MIKLFQKIRYHWQPELSWSIIYWSIAFAPIFVGLSLLYERTEIPSRVFILFAIFAVLVGI 60

25 Query: 61 GFHRYFVIEEDGYLRIVSFNFLRRTKFPIEDIAKIEVTKSSVTIKFNNNHERIFYMRKWP 120
G HRYF+IE +G LRIVSF K I I KIEVTKS++ + + +FYM RKWP
Sbjct: 61 GLHRYFIIENNGILRIVSFKLFGRKLLISTITKIEVTKSTLCLHVEDK-SYLFYMRKWP 119

Query: 121 KKYFLDALAIEPTFKGEVELLDNLIKMDYFECYRYDKKALTK 162
KKYFLDALA+ P F+GEV L DN IK+DYFE Y++DKKALT+

30 Sbjct: 120 KKYFLDALAVNPYFQGEVILSDNFIKLDYFEVYQHDKKALTR 161

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1181

35 A DNA sequence (GBSx1257) was identified in *S.agalactiae* <SEQ ID 3675> which encodes the amino acid sequence <SEQ ID 3676>. This protein is predicted to be peptidase t (pepT). Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA20627 GB:L27596 tripeptidase [Lactococcus lactis]
Identities = 274/406 (67%), Positives = 334/406 (81%), Gaps = 4/406 (0%)

50 Query: 1 MSYEKLLERFLTYVKINTRSNPNSTQPTTQSQVDFALTVLKPEMEAIGLKD VHYLPSNG 60
M YEKLL RFL YVK+NTRS+ NST TP+TQ+ V+FA + +M+A+GLKD VHYL SNG
Sbjct: 1 MKYEKLLPRFLEYVKVNTRSDENSTTTPSTQALVEFAHK-MGEDMKALGLKD VHYLESNG 59

55 Query: 61 YLVGTLPATSDRLRHKIGFISHMDTADFNAENITPQIVDYKGGD--IELGDSGYILSPKD 118
Y++GT+PA +D+ KIG ++H+DTADFNAE + PQI++ G+ I+LGD+ + L PKD
Sbjct: 60 YVIGTIPANTDKKVRKIGLLAHLDTADFNAEGVNPQILENYDGESVIQLGDTEFTLDPKD 119

Query: 119 FPNLNHYHGQTLITTDGKTLTGADDKSGIAEIMTAMEYLAS-HPETEHCEIRVGFPGPDEE 177
FPNL NY GQTL+ TDG TLLG+DDKSG+AEIMT +YL + +P+ EH EIRVGFPGPDEE

-1324-

Sbjct: 120 FPNLKNYKGQTLVHTDGTTLGSDDKSGVAEIMTLADYLLNINPDEFHGEIRVGFPGDEE 179

Query: 178 IGIGADKFDVKDFDVFAYTVDGGPLGELQYETFSAGLELTFEGRNVHPGTAKNQMINA 237
IG+GADKFDV DFDVDFAYTVDGGPLGELQYETFSAG + F+G+NVHPGTAKN M+NA

5 Sbjct: 180 IGVGADKFDVADFDVDFAYTVDGGPLGELQYETFSAGAVIEFQGNVHPGTAKNMMVNA 239

Query: 238 LQLAMDFHSQLPENERPEQTDGYQGIFYHLYDLSGTVDQAKSSYIIRDFEEVDFLKRKHLA 297
LQLA+D+H+ LPE +RPE+T+G +GF+HL L GT ++A++ YIIRD EE F +RK L

10 Sbjct: 240 LQLAIDYHNALPEFDRPEKTEGREGFFHLKLDGTPEEARAQYIIRDHEEGKFNERKALM 299

Query: 298 QDIADNMNEALQSERVKVLYDQYYNMKKVIEKDMTPINIAKEVMEELDIPPIEPIRG 357
Q+IAD MN L RVK + DQYYNM ++IEKDM+ I+IAK+ ME LDI PIIEPIRG

Sbjct: 300 QEIADKMNAELGQNRVKPVIKDQYYNMAQIIEKDMSTIIDIAKKAMENLDIAPPIEPIRG 359

15 Query: 358 TDGSKISFMGIPTPNLFAGGENMHGRFEFVSLQTMKAVDVILGIV 403
TDGSKISFMG+PTPNLFAGGENMHGRFEFVS+QTMKAVD +L I+

Sbjct: 360 TDGSKISFMGLPTPNLFAGGENMHGRFEFVSVQTMKAVDTLLEII 405

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3677> which encodes the amino acid
20 sequence <SEQ ID 3678>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2938(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 305/406 (75%), Positives = 352/406 (86%), Gaps = 1/406 (0%)

Query: 1 MSYEKLLERFLTYVKINTRSNPNSTQPTTQSQVDFALTVLKPEMEAIGLKD VHYLPSNG 60
M Y+ LL+RF+ YVK+NTRS P+S TP+T+SQ FALT+LKPEMEAIGL+DVHY P NG

35 Sbjct: 5 MKYDNLDRFIKYVKVNTSRVPDSETPSTESQEA FALTILKPEMEAIGLQDVHYNPVG 64

Query: 61 YLVGTLPATSDRLRHKIGFISHMDTADFNAENITPQIVD-YKGGDIELGDSGYILSPKDF 119
YL+GTLPA + L KIGFI+HMDTADFNAEN+ PQI+D Y+GGDI LG S Y L PK F

Sbjct: 65 YLIGTLPANNPTLTKIGFIAHMDTADFNAENVNPQIIDNYQGGDITLGSSNYKLDPKAF 124

40 Query: 120 PNLNNYHGQTLITTDGKTLGADDKSGIAEIMTAMEYLASHPEIEHCEIRVGFPGDEEIG 179
PNLNNY GQTLITTDG TLLGADDKSGIAEIMTA+E+L S P+IEHC+I+V FGPDEEIG

Sbjct: 125 PNLNNYIGQTLITTDGTTLLGADDKSGIAEIMTAIEFLTSQPQIEHCDIKVAFGPDEEIG 184

45 Query: 180 IGADKFDVKDFDVFAYTVDGGPLGELQYETFSAGLELTFEGRNVHPGTAKNQMINALQ 239
+GADKF+V DF+VDFAYT+DGGPLGELQYETFSAA LE+TF GRNVHPGTAK+QMINAL+

Sbjct: 185 VGADKFEVADFEVDFAITMDGGPLGELQYETFSAAALEVTFLGRNVHPGTAKDQMINALE 244

50 Query: 240 LAMDFHSQLPENERPEQTDGYQGIFYHLYDLSGTVDQAKSSYIIRDFEEVDFLKRKHLAQD 299
LA+DFH +LP +RPE TDGYQGIFYHL L+GTV++A++SYIIRDFEE F RK ++

Sbjct: 245 LAIDFHEKLPKDRPEYTDGYQGIFYHLTGLTGTVEEARASYIIRDFEASFEARKVKVEN 304

55 Query: 300 IADNMNEALQSERVKVLYDQYYNMKKVIEKDMTPINIAKEVMEELDIPPIEPIRG 359
IA +MN L ++RV V+L DQYYNMKKVIEKDMT I +AKEVMEEL IKP+IEPIRG

Sbjct: 305 IAQSMNAQLGTRKRVLVELNDQYYNMKKVIEKDMTAEI LAKEVMEELAIKPVIEPIRG 364

Query: 360 GSKISFMGIPTPNLFAGGENMHGRFEFVSLQTMKAVDVILGIVAK 405
GSKISFMGIPTPN+FAAGGENMHGRFEFVSLQTMK+AVDVI+G+V K

Sbjct: 365 GSKISFMGIPTPNIFAGGENMHGRFEFVSLQTMERAVDVIIGLVCK 410

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

-1325-

Example 1182

A DNA sequence (GBSx1258) was identified in *S.agalactiae* <SEQ ID 3679> which encodes the amino acid sequence <SEQ ID 3680>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood ==-12.26    Transmembrane  481 - 497 ( 477 - 508)
    INTEGRAL    Likelihood = -9.45     Transmembrane  510 - 526 ( 506 - 534)
    INTEGRAL    Likelihood = -7.96     Transmembrane  316 - 332 ( 310 - 334)
    INTEGRAL    Likelihood = -7.54     Transmembrane  354 - 370 ( 351 - 373)
10  INTEGRAL    Likelihood = -7.11     Transmembrane  385 - 401 ( 383 - 409)
    INTEGRAL    Likelihood = -6.58     Transmembrane  215 - 231 ( 211 - 233)
    INTEGRAL    Likelihood = -6.48     Transmembrane   71 -  87 (  69 -  91)
    INTEGRAL    Likelihood = -6.32     Transmembrane  110 - 126 ( 106 - 133)
    INTEGRAL    Likelihood = -5.10     Transmembrane  446 - 462 ( 443 - 465)
15  INTEGRAL    Likelihood = -3.29     Transmembrane  418 - 434 ( 418 - 435)
    INTEGRAL    Likelihood = -2.55     Transmembrane  263 - 279 ( 263 - 279)
    INTEGRAL    Likelihood = -2.02     Transmembrane  142 - 158 ( 141 - 159)
    INTEGRAL    Likelihood = -1.70     Transmembrane  184 - 200 ( 184 - 200)

20  ----- Final Results -----
        bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
        bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

25 A related GBS nucleic acid sequence <SEQ ID 8747> which encodes amino acid sequence <SEQ ID 8748> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 6
McG: Discrim Score:      -10.58
GvH: Signal Score (-7.5): -1.1
30  Possible site: 32
    >>> Seems to have no N-terminal signal sequence
ALOM program    count: 13 value: -12.26 threshold:  0.0
    INTEGRAL    Likelihood ==-12.26    Transmembrane  470 - 486 ( 466 - 497)
    INTEGRAL    Likelihood = -9.45     Transmembrane  499 - 515 ( 495 - 523)
35  INTEGRAL    Likelihood = -7.96     Transmembrane  305 - 321 ( 299 - 323)
    INTEGRAL    Likelihood = -7.54     Transmembrane  343 - 359 ( 340 - 362)
    INTEGRAL    Likelihood = -7.11     Transmembrane  374 - 390 ( 372 - 398)
    INTEGRAL    Likelihood = -6.58     Transmembrane  204 - 220 ( 200 - 222)
    INTEGRAL    Likelihood = -6.48     Transmembrane   60 -  76 (  58 -  80)
40  INTEGRAL    Likelihood = -6.32     Transmembrane   99 - 115 (  95 - 122)
    INTEGRAL    Likelihood = -5.10     Transmembrane  435 - 451 ( 432 - 454)
    INTEGRAL    Likelihood = -3.29     Transmembrane  407 - 423 ( 407 - 424)
    INTEGRAL    Likelihood = -2.55     Transmembrane  252 - 268 ( 252 - 268)
    INTEGRAL    Likelihood = -2.02     Transmembrane  131 - 147 ( 130 - 148)
45  INTEGRAL    Likelihood = -1.70     Transmembrane  173 - 189 ( 173 - 189)
    PERIPHERAL  Likelihood =  1.43      21
    modified ALOM score:  2.95

*** Reasoning Step: 3
50  ----- Final Results -----
        bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
        bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
55

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC00276 GB:AF008220 YtgP [Bacillus subtilis]
Identities = 178/545 (32%), Positives = 302/545 (54%), Gaps = 26/545 (4%)

60  Query: 24  QMVKGTAWLTAGNFISRLLGAIYIIPWYAWMGKHAEEANALFGMGYEIYALFLLISTVGI 83
      ++++GT  LT G +ISR+LG +Y+IP+  +G  A  ALF  GY  Y LFL I+T+G
Sbjct: 4     KLLRGTFVLTGLGTYISRLIGMVYLIPFSIMVG---ATGGALFQYGYNQYTLFLNIATMGF 60

```

-1326-

Query: 84 PVA VAKQVSKYNTLGKEEMSTIYLVRKILQFMLILGGIFALIMYIGSPLFASLSKGGQE-- 141
 P AV+K VSKYN+ G E S +++ + ML+ G I I+Y+ +P+FA +S GG++
 Sbjct: 61 PAAVSKFVSKYNSKGDYETSRKMLKAGMSVMLVTGMIAFFILYLSAPMFAEISLGGKDN 120

5 Query: 142 -----LVPILRSLTLAVLVFSPMSVLRGFFQGFNNLKPYAISQVAEQIIRVIWMLLTAF 195
 +V ++R ++LA+LV P MS++RGFFQG + P A+SQV EQI+R+I++L F
 Sbjct: 121 GLTIDHVVVYVIRMVSLALLVVPIMSLVRGFFQGHQMMGPTAVSQVVEQIVRIIFLLSATF 180

10 Query: 196 YIMRLGSGDYIAAVTQSTFAAFVGMFASIAVLLYFLW--RYNMLSALIGKTPKHIKLDTK 253
 I+++ +G + AV +TFAA +G F + V+LY W R L A++ T L K
 Sbjct: 181 LILKVFNGGLVIAVGATFAALIGAFGGL-VVLYIYWNKRKGSLLAMPNTGTANLSYK 239

15 Query: 254 EILIIETIKEAIPFIITGAAIQIFKLIDQFSFGNTM--ALFTNYSSEELRVMFAYFSSNPG 311
 ++ E A P++ G AI ++ ID +F M A S + L ++ Y
 Sbjct: 240 KMFFELFYSYAAPYFVGLAIPLYNYIDTNTFNKAMIEAGHQAI SQDMLAILTLVQ---- 295

20 Query: 312 KVTMILIAVATAIAGVGIPLLTENFVKNDKKAARLVVNNLQMLLMFLLPAVAGSVILAK 371
 K+ MI +++ATA IP +TE+F + K + + +Q +L ++PAV G +L+
 Sbjct: 296 KLVMI PVSLATAFGLTLIPTITESFTSGNYKLLNQINQTMQTILFLIIPAVVGISLLSG 355

Query: 372 PLYTVFYGL----PQGQALGLFVISLIQTIILSIYTVLAPMLQALFENRKAIYFLYGLV 427
 P YT FYG P+ A L S + I+ S++TV A +LQ + + + A++ + G+V
 Sbjct: 356 PTYTFYFYGSESLHPELGANILLWYSPV-AILESLFTVNAAILQGINQKQFAVVS LVIGVV 414

25 Query: 428 AKVILQLPSIFLFHAYGPLFSTTVALCIPVILMYLKIHEITGFKRQAIRRTSALVLILTL 487
 K++L +P I L A G + +T + ++ ++ I G+ + + + + L+L+L+
 Sbjct: 415 IKLVNLVPLIKMQADGAILATALGYIASLLYGFIMIKRHAGYSYKILVKRTVLMVLVLSA 474

30 Query: 488 LMSFIISMIIWLMNLVI-VPDSRLVSLVYIIVIGAIGLVYGFMALATHLLDKMIGSRAQ 546
 +M + ++ W++ I D ++ + + +++ A+G VY + L K++G R
 Sbjct: 475 IMGIAVKIVQWVLGFFISYQDGMQAAIVVVIAAAVGGAVYLYCGYRLGFLQKILGRRLP 534

Query: 547 DLRRK 551
 RK
 35 Sbjct: 535 GFFRK 539

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3681> which encodes the amino acid sequence <SEQ ID 3682>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.60	Transmembrane	468 - 484 (466 - 493)
INTEGRAL	Likelihood = -8.39	Transmembrane	305 - 321 (299 - 323)
INTEGRAL	Likelihood = -7.75	Transmembrane	343 - 359 (340 - 362)
INTEGRAL	Likelihood = -6.58	Transmembrane	374 - 390 (373 - 398)
INTEGRAL	Likelihood = -4.25	Transmembrane	138 - 154 (137 - 157)
INTEGRAL	Likelihood = -3.45	Transmembrane	100 - 116 (98 - 122)
INTEGRAL	Likelihood = -3.40	Transmembrane	415 - 431 (410 - 432)
INTEGRAL	Likelihood = -3.35	Transmembrane	499 - 515 (499 - 519)
INTEGRAL	Likelihood = -2.60	Transmembrane	433 - 449 (432 - 451)
INTEGRAL	Likelihood = -2.50	Transmembrane	173 - 189 (173 - 190)
INTEGRAL	Likelihood = -0.59	Transmembrane	201 - 217 (201 - 220)

----- Final Results -----

bacterial membrane	--- Certainty=0.4439(Affirmative) < succ>
bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC00276 GB:AF008220 YtgP [Bacillus subtilis]
 Identities = 169/536 (31%), Positives = 295/536 (54%), Gaps = 24/536 (4%)

Query: 14 MVQGAAWSTAGNFISRLGLVLYIIPWYIWMGQYAIQANALFNMGYNVYAYFLLISTTGLN 73
 +++G T G +ISR+LG++Y+IP+ I +G ALF GYN Y FL I+T G
 Sbjct: 5 LLRGTFVLTGLTYISRLGMVYLIPFSIMVGA---TGGALFQYGYNQYTLFLNIATMGFP 61

-1327-

Query: 74 VAIKQVAKYNSMGQTEHSYQLIRSTLKLMLGLGLIFSAIMYLGSPLEFASLS-GGDDT-- 130
 A++K V+KYNS G E S +++++ + +ML G+I I+YL +P+FA +S GG D
 Sbjct: 62 AAVSKFVSKYNSKGDYETSRKMLKAGMSVMLVTGMTAFFILYLSAPMFAEISLGKDNNG 121

5 Query: 131 -----LVPIMHSLSLAVFIFPVMSVIRGIFQGHNNIKPYAVSQIAEQILIRVIWMLLTTF 185
 +V ++ +SLA+ + P+MS++RG FQGH + P AVSQ+ EQ++R+I++L TF
 Sbjct: 122 LTIDHVYVIRMVSLALLVVPIMSLVRGFFQGHQMMGPTAVSQVVEQIVRIIFLLSATFL 181

10 Query: 186 IMKLGSGDYASAVTQSTFAAFIGMVASMGVLGYLW--KQGLLAIFSKPDHTVSDIK 243
 I+K+ +G AV +TFAA IG + VL Y W ++G L A+ T ++ K
 Sbjct: 182 ILKVFNGGLVIAVGATFAALIGAFGLVLV-YIYWNKRKGSLLAMPNTGPTANLSYKK 240

15 Query: 244 LLETLKESIPFIVTGSIAIQAFQLIDQWTFVNTMTLFTDYSRSQ--LLVLFYGFNANPAK 301
 + E + P++ G AI + ID TF M + SQ L +L Y K
 Sbjct: 241 MFFELFSYAAPYVFGLAIPLYNYIDTNTFNKAMIEAGHQAIQDMLAILTLYVQ----K 296

20 Query: 302 ITMVLIABAASIGGVGIALLTENYVKKDKMAAARLIINNIEMLMVFLPALTGAILARP 361
 + M+ +++A + G I +TE++ + K + I ++ ++ ++PA+ G +L+ P
 Sbjct: 297 LVMIPVSLATAFGLTLIPTITESFTSGNYKLLNQINQTMQTILFLIIPAVVGISLLSGP 356

25 Query: 362 LYSVFGYASE---ERAIHLFVAVLFQTLALLALYTLFSPMLQALFENRKAIYYFAYGILIK 418
 Y+ FYG+ E ++ + +L +L+T+ + +LQ + + + A+ G++IK
 Sbjct: 357 TYTFFYGSSESLHPELGANILLWYSPVAILFSLFTVNAAILQGINKQKFAVSVLVIGVVIK 416

30 Query: 419 LVLQIPLIYLLHAYGPLLATTIALVVPYIYLMYRRLYQVTHFNKLLQKRLLLTLIETLLM 478
 LVL +PLI L+ A G +LAT + + + + + ++ K+L KR +L L+ + +M
 Sbjct: 417 LVLNVPLIKLMQADGAILATALGYIASLLYGFIMIKRHAGYSYKILVKRVTMLVLSAIM 476

Query: 479 GLVVVFVANWLLGYAFK-PTGRLTSLYLLLIIGGLGMTVYTALTLTHQDKLIGSK 533
 G+ V + W+LG+ G++ + + ++I +G VY L K++G +
 Sbjct: 477 GIAVKIVQWVLGFFISYQDGMQAAIVVVIAAAGGAVYLYCGYRLGFLQKILGRR 532

An alignment of the GAS and GBS proteins is shown below.

Identities = 320/541 (59%), Positives = 431/541 (79%)

35 Query: 12 MSQKTKVKSQQEQMVKGTAWLTAGNFISRLGAIYIIPWYAWMGKHAEEANALFGMGYEI 71
 MS + +++Q+E MV+G AW TAGNFISRLG +YIIPWY WMG++A +ANALF MGY +
 Sbjct: 1 MSTKKQLTQEELMVQGAAWSTAGNFISRLGLVLYIIPWYIWMGQYAIQANALFNMGYNV 60

40 Query: 72 YALFLLISTVGIPVAVAKQVSKYNTLGKEMSIYLVKILQFMLILGGIFALIMYIGSPL 131
 YA FLLIST G+ VA+AKQV+KYN++G+ E S L+R L+ ML LG IF+ IMY+GSPL
 Sbjct: 61 YAYFLLISTTGLNVAIAKQVAKYNSMGQTEHSYQLIRSTLKLMLGLGLIFSAIMYLGSPLE 120

45 Query: 132 FASLSKGGQELVPILRSITLAVLVFSPMSVLRGFFQGFNNLKPYAISQVABEQIIRVIWML 191
 FASLS G LVPI+ SL+LAV +FP MSV+RG FQG NN+KPYA+SQ+AEQ+IRVIWML
 Sbjct: 121 FASLSGGDDTLVPIMHSLSLAVFIFPVMSVIRGIFQGHNNIKPYAVSQIAEQILIRVIWML 180

50 Query: 192 LTAFYIMRLGSGDYIAAVTQSTFAAFVGMFASIAVLLYFLWRYNMLSALIGKTPKHIKLD 251
 LT F+IM+LGSGDY +AVTQSTFAAF+GM AS+ VL Y+LW+ +L+A+ K + +D
 Sbjct: 181 LTTFFIMKLGSGDYASAVTQSTFAAFIGMVASMGVLGYLWQGLLAIFSKPDHTVSD 240

55 Query: 252 TKEILTIETIKEAIPFIITGAAIQIFKLIDQFSFGNTMALFTNYSSEELRVMFAYFSSNPG 311
 K +L+ET+KE+IPFI+TG+AIQ F+LIDQ++F NTM LFT+YS +L V+F YF++NP
 Sbjct: 241 IKGLLLETLKESIPFIVTGSIAIQAFQLIDQWTFVNTMTLFTDYSRSQLLVLFYGFNANPA 300

60 Query: 312 KVTMILIAVATAIAGVGIPLLTENFVKNDKKAARLVVNNLQMLLMFLLPVAVAGSVILAK 371
 K+TM+LIAVA +I GVGI LLTEN+VK D KAAARL++NN++ML+MFLLP+ G++ILA+
 Sbjct: 301 KITMVLIABAASIGGVGIALLTENYVKKDKMAAARLIINNIEMLMVFLPALTGAILAR 360

65 Query: 372 PLYTVFYGLPQGGALGLFVISLIQTIILSIYTVLAPMLQALFENRKAIITYFLYGLVAKVI 431
 PLY+VFYG + +A+ LFV L QT++L++YT+ +PMLQALFENRKAI YF YG++ K++
 Sbjct: 361 PLYSVFYGASERAIHLFVAVLFQTLALLALYTLFSPMLQALFENRKAIITYFAYGILIKLV 420

Query: 432 LQLPSIFLFHAYGPLFSTTVALCIPVILMYLKIHEITGFKRQAIRTSALVLILTLLMSF 491
 LQ+P I+L HAYGPL +TT+AL +P+ LMY +++++T F R+ +++ L LI TLLM
 Sbjct: 421 LQIPLIYLLHAYGPLLATTIALVVPYIYLMYRRLYQVTHFNKLLQKRLLLTLIETLLMGL 480

-1328-

Query: 492 IISMIIWLMNLVIVPDSRLVSLVYIIVIGAIGLVYGFMAIATHLLDKMIGSRAQDLRRKL 552
 ++ + WL+ P RL SL+Y+++IG +G+ VY + L TH LDK+IGS+A LR+KL
 Sbjct: 481 VVFVANWLLGYAFKPTGRLTSLLYLLIIGGLGMTVYTALTLTHQLDKLIGSKASRLRQKL 541

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1183

A DNA sequence (GBSx1259) was identified in *S.agalactiae* <SEQ ID 3683> which encodes the amino acid sequence <SEQ ID 3684>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4104(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06290 GB:AP001515 UDP-N-acetylmuramoylalanyl-D-glutamyl-2,
 6-diaminopimelate ligase [Bacillus halodurans]
 Identities = 153/468 (32%), Positives = 237/468 (49%), Gaps = 23/468 (4%)

Query: 33 NVTFNALSYSRQISSDTLFFA-KGATFK-KEYLDSAITAGLSFYVSETDYGADIPVILV 90
 N +++ DSR++ LFF KG T +Y A++ G VSE +PV++V
 Sbjct: 21 NPDHSHSMDSREVVEGGLFFCIKGYTVGDHDAQQAVSNGAVVVSERPLELSVPVVVV 80

Query: 91 NDIKKAMSLISMSFYNNPQNKLLAFTGTGKTTAAYFAYHMLKVNHR-PAMLSTMNTT 149
 D ++AM+ ++ FY P N L+L+ TGT GKTT + +++ + ++ TM T
 Sbjct: 81 RDSRRAMAQVATKFYGEPTNDLQLIGVTGTNGKTTITHLIEKIMQDQGKMTGLIGTMYTK 140

Query: 150 LDGKSFFKSHLTPESLDLFRMMATAVENQMTHLIMEVSSQAYLTKRVYGLTFDVGVFILN 209
 + G ++ TTPESL L R A ++ +T +MEVSS A + RV G FDV VF N
 Sbjct: 141 I-GHEHKETKNTPESLVLQRTFADMKKSGVTTAMMEVSSHALQSGRVRGCDFDVAVFSN 199

Query: 210 ISPDHIGPIEHPTEFEDYFFHKRLME-----NSNAVNVN-----SQMDHFNIVKEQVEYI 259
 ++PDH+ H T E Y F K LL V+N + D + QV
 Sbjct: 200 LTPDHLDD--YHGTMYRYKFAKGLLFAQLGNTYQGVAVLNADDPASADFAEMTIAQVVITY 257

Query: 260 PHDFYGDY-SENVITESKAFSPHVKGKLEN-TYDIKLIKFNQENAIAGLACLRGLVSI 317
 + D+ +ENV S +F + E I LIGKF+ N +AA A GV +
 Sbjct: 258 GIENEADFQAENVRITSTGTTFFELAAFEERMELSIHLIGKFSVYNVLAATAAAAYVSGVPL 317

Query: 318 EDIKNGIAQTT-VPGRMEVLTQTNGAKIFVDYAHNGDSLKLLAVVEEHQKGDIIILVLA 376
 ++IK + + V GR E + + VDYAH DSL+ +L V E KGD+ +V+G
 Sbjct: 318 QEIKKSLEEVKGVAGRFFETVKHDQPFVTVVDYAHTPDLENVLKTVGELAKGDVRRVVVGC 377

Query: 377 PGNGKQSRKDFGVDVNHQPNLQVILTADDPNFEDPLVISQEIASHINRPVTIII-DREE 435
 G++ +++R ++ N Q I T+D+P E+P+ I +++ ++I DR+E
 Sbjct: 378 GGDRDKTKRPVMAEIAITTFAN-QAIFTSDNPRSEEPMDILRDMEQGAKGDSYLMIEDRKE 436

Query: 436 AIANASTLTNCKLDALIIAGKGADAYQIIKGNRDNYSGDLEVAKKYLK 483
 AI A L + D I+IAGKG + YQ + ++ D VA++ +K
 Sbjct: 437 AIFKAIEIAK-EDDIIVIAKGHETYQQFRDRTIDFD-DRIVAQQAIAK 482

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3685> which encodes the amino acid sequence <SEQ ID 3686>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1329-

bacterial cytoplasm --- Certainty=0.4717(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 350/482 (72%), Positives = 399/482 (82%), Gaps = 1/482 (0%)

Query: 1 MITIDKILEILKNDHNFREILFHEHYNNWTQNVTFNALSYSRQISSDTLFFAKGATFK 60
 MITI+++L+ILK DHNFRE+L + Y+Y++ Q +F LSYSRQ+ TLFFAKGATFK
 10 Sbjct: 1 MITIEQLLDILKKDHNPREVLADGYHYHY-QGFSFERLSYSRQVDGKTLFFAKGATFK 59

Query: 61 KEYLDSAITAGLSFYVSETDYGADIPVILVNDIKKAMSLISMSFYNNPQNKLLAFTGT 120
 +YL AIT GL Y+SE DY IPV+LV DIKKAMSLI+M+FY NPQ KLKLLAFTGT
 15 Sbjct: 60 ADYLKEAITNGLQLYISEVDYELGIPVVLVTDIKKAMSLIAMAFYGNPQEKLLAFTGT 119

Query: 121 KGKTTAAYFAYHMLKVNHRPAMLSMTNTTLDGKSFFKSHLTPESLDLFRMMATAVENQM 180
 KGKTTAAYFAYHMLK +++PAM STMTNTLDGK+FFKS LTPESLDLF MMA V N M
 20 Sbjct: 120 KGKTTAAYFAYHMLKESYKPMFSTMTNTLDGKTFFKSOLTPESLDLFAMMAECVTNGM 179

Query: 181 THLIMEVSSQAYLTKRVYGLTFDVGVLNISPDIHIGPIEHPTFEDYFFHKRLLMENSNAV 240
 THLIMEVSSQAYL RVYGLTFDVGVLNISPDIHIGPIEHPTFEDYF+HKRLLMENS AV
 25 Sbjct: 180 THLIMEVSSQAYLVDRVYGLTFDVGVLNISPDIHIGPIEHPTFEDYFYHKRLLMENSRAV 239

Query: 241 VVNSQMDHFNIVKEQVEYIPHDYFGDYSENVITESKAFSFBVKGKLENTYDIKLIKGFNQ 300
 V+NS MDHF+ + +QV H FYG S+N IT S+AFSF KG+L YDI+LIG FNQ
 30 Sbjct: 240 VINSGMDHFSFLADQVADQEHVFGPLSDNQITTSQAFSFEAKGQLAGHYDIQLIGHFNQ 299

Query: 301 ENAIAAGLACLRGLGVSIEDIKNGIAQTTPVGRMEVLTQTNGAKIFVDYAHNGDSLKKLLA 360
 ENA+AAGLACLRGLG S+ DI+ GIA+T VPGRMEVLT TN AK+FVDYAHNGDSL+KLL+
 35 Sbjct: 300 ENAMAAGLACLRGLGASLADIQKGIKTRVPGRMEVLTMTNHAKVFDYAHNGDSLEKLLS 359

Query: 361 VVEEHQKGDIIILVLAGPKNKGQSRKDFGDIVNQHNPQLVILTADDPNFEDPLVISQEI 420
 VVEEHQ G ++L+LGAPKNKG+SRR DFG VI+QHPNL VILTADDPNFEDP IS+EIA
 40 Sbjct: 360 VVEEHQTGKMLLILGAPKNKGESRRADFGRVINQHNPQLTVILTADDPNFEDPEDISKEIA 419

Query: 421 SHINRPVTIIIDREEAIAANASTLTNCKLDIAIIAGKGADAYQIIKGNRDNYSGDLEVAKKYL 482
 SHI RPV II DRE+AI A +L DA+IIAGKGADAYQI+KG + Y+GDL +AK YL
 45 Sbjct: 420 SHIARPVEIISDREQAIQKAMSLCQGAQDAVITAGKGADAYQIVKGQVAYAGDLAIKHYL 481

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1184

A DNA sequence (GBSx1260) was identified in *S.agalactiae* <SEQ ID 3687> which encodes the amino acid sequence <SEQ ID 3688>. Analysis of this protein sequence reveals the following:

45 Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1330-

Example 1185

A DNA sequence (GBSx1261) was identified in *S.agalactiae* <SEQ ID 3689> which encodes the amino acid sequence <SEQ ID 3690>. This protein is predicted to be FhuA (fepC). Analysis of this protein sequence reveals the following:

5 Possible site: 54
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2785(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9975> which encodes amino acid sequence <SEQ ID 9976> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

 >GP:AAF98153 GB:AF251216 FhuC [Staphylococcus aureus]
 Identities = 141/259 (54%), Positives = 193/259 (74%)

20 Query: 7 MSHIKAENIIVSYDQKEIINNLSLSILNQKITTIIGANGCGKSTLLKALTRIHKIKDGTI 66
 M+ + + + Y IIN L + I + K+T+IIG NGCGKSTLLKAL+R+ +K+G +
 Sbjct: 1 MNRLHGGQVVKIGYGDNTIINKLDVEIPDGKVTSIIGPNGCGKSTLLKALSRLAVKEGEV 60

 Query: 67 TIDGHDIAHLPTKEIAKKIALLPQVLEATEGITVYELISYGRFPHQKYLGNLTNDDRSKI 126
 +DG +I TKEIAKKIA+LPQ E +G+TV EL+SYGRFPHQK G LT +D+ +I
 25 Sbjct: 61 FLDGENIHTQSTKEIAKKIAILPQSPEVADGLTVGELVSYGRFPHQKGFGRLTAEKKEI 120

 Query: 127 HWAMEMTNVAQFANRDVDDLGGQRQKVWIAMALAQDITDIFLDEPTTYLDMNHQLEVLE 186
 WAME+T F +R ++DLGGQRQ+VWIAMALAQ TD IFLDEPTTYLD+ HQLE+LE
 30 Sbjct: 121 DWAMEVTGTDTFRHRSINDLSGGQRQVRWIAMALAQRTDIIIFLDEPTTYLDICHQLEILE 180

 Query: 187 LLKKLNDETQKTIIMVLHDLNLSARYSDYLVAMKTGKIIYEGSPSQIMTKDIIDIFKID 246
 L++KLN E TI+MVLHD+N + R+SD+L+AMK G II GS ++T++I++ +F ID
 Sbjct: 181 LVQQLNQEQGCTIVMVLHDINQAIRFSDDLHIAKMGDIIATGSTEDVLTQEILEKVFNID 240

35 Query: 247 AHIIQDPISKQPVLLSYQL 265
 + +DP + +P+L++Y L
 Sbjct: 241 VVLKDPKPTGKPLLVTYDL 259

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1929> which encodes the amino acid sequence <SEQ ID 1930>. Analysis of this protein sequence reveals the following:

 Possible site: 48
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2970(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 166/259 (64%), Positives = 208/259 (80%)

 Query: 7 MSHIKAENIIVSYDQKEIINNLSLSILNQKITTIIGANGCGKSTLLKALTRIHKIKDGTI 66
 M+ I AE++ ++Y+Q+ II+ LS I KITTIIGANGCGKS+LLKALTR+ K G +
 Sbjct: 1 MTTISAEDLTIAEQRTIIDKLSFYIPEGKITTIIGANGCGKSSLLKALTRLLPPKQGVV 60

55 Query: 67 TIDGHDIAHLPTKEIAKKIALLPQVLEATEGITVYELISYGRFPHQKYLGNLTNDDRSKI 126
 ++G +IA L TKE+AKK+ALLPQV EAT GITVYEL+SYGRFPHQ Y GNL+ D+ I
 Sbjct: 61 YLNGQNIATLETKEVAKKLALLPQVQEATNGITVYELVSYGRFPHQSYFGNLSPADKKAI 120

-1331-

Query: 127 HWAMEMTNVAQFANRDVDDLSSGGQRQKQVWIAMALAQDTDTIFLDEPTTYLDMNHQLEVL 186
 HWAM+ TNV +A++ VD LSGGQRQ+VW+AMALAQ TDTIFLDEPTTYLD+NHQLE+LE
 Sbjct: 121 HWAMQATNVMAQADQPDALSSGGQRQKQVWILAMALAQDTDTIFLDEPTTYLDLNHQLLEILE 180

5 Query: 187 LLKKLNDETQKTIIMVLHDLNLSARYSDYLVAMKTGKIIEGSPSQIMTKDIIKDIFKID 246
 L+K LN + KTI+MVLHDLNLSARYSD+L+AMK GKI Y G+ + +MT II+DIF+I
 Sbjct: 181 LVKSLNKDAGKTIVMVLHDLNLSARYSDHLIAMKHGKIHYTGTIADVMTSPIIQDIFQIK 240

10 Query: 247 AHIIQDPISKQPVLLSYQL 265
 ++ DPI P++L+YQL
 Sbjct: 241 PVLVDDPIHNCPVLTYQL 259

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1186

A DNA sequence (GBSx1262) was identified in *S.agalactiae* <SEQ ID 3691> which encodes the amino acid sequence <SEQ ID 3692>. This protein is predicted to be ferrichrome ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07609 GB:AP001520 ferrichrome ABC transporter
 (ferrichrome-binding protein) [Bacillus halodurans]
 Identities = 94/301 (31%), Positives = 177/301 (58%), Gaps = 11/301 (3%)

30 Query: 6 IIVLTLLTFFLV---SCGQQTQKQESTKTTISK--MPKIEGFTYYGKIPENPKKVINFYTS 60
 +++LT+L F L+ +CG T E S+ M E T ++P NP++V+
 Sbjct: 7 LLLLTMLLFALLVVAACGSNTDAEQADELESEDMITYESETGPIEVFANPQRVV--ALG 64

35 Query: 61 YTGYYLLKLGNNVSSYSLEKDSPVFGKQKEAKKLTAADTEAIAAQKPDLMVFDQDPN 120
 +TG +L L VNV K++P + + L++ +++ ++ E I PDLI+ + N
 Sbjct: 65 FTGNILALDVNVVGVDL-WSKNNPNYEQLLQDVTVEVSEENLEQIMELDPDLIIAYSTVQN 123

40 Query: 121 INTLKKIAPTLVIKYGAQNYLDMMPALGKVFGEKEANQWVSQWKTTLAVKKDLHHILK 180
 L++IAPT++ Y +YL+ +GK+ KE+EA WV +K + +++ +
 Sbjct: 124 AEQLQEIAPTFLVLYTNNLDYLEQHVIGKLLNKEEAQAWVDDFKARAEQAGEEIKEKIG 183

45 Query: 181 PNTTFTIMDFYDKNIYLYGNNFGRGGELIYDSLGYAAPEKVKKDVFKKGWFTVSQEAIGD 240
 + T ++++ ++ +Y++GNN+GRG E++Y ++ A PE+V++ G++ +S EA+ +
 Sbjct: 184 EDATVSVIETFEDQLYVFGNNWGRGTEILYQTMDLAMPERVEEMALADGYIALSFEALPE 243

50 Query: 241 YVG DYALVNINKTTPKAASSLKESDVWKNLPAVKKGHIIESNYDVVFYFSDPLSLEAQLKSF 30~
 + GDY +++ N +A +S +E++ ++++PAV+ G + E+N FYF+DPLSLE QL+ F
 Sbjct: 244 FAGDYIILSKN---DEADNSFQETINTYQSIPAVQNGQVFANAEKIFYFNDPLSLELQLEFF 301

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3693> which encodes the amino acid sequence <SEQ ID 3694>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> May be a lipoprotein

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1332-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:BAB07609 GB:AP001520 ferrichrome ABC transporter
 (ferrichrome-binding protein) [Bacillus halodurans]
 Identities = 112/306 (36%), Positives = 178/306 (57%), Gaps = 3/306 (0%)

10 Query: 2 KKLTLTLTCLTTITLIACGNQATNHSNTASKSLSPMPQIAGVTYYGDIPKQPKRVVSLA 61
 K L LL L + + ACG+ +S M T ++P P+RVV+L
 Sbjct: 5 KHLTLTLLTCLTTITLIACGNQATNHSNTASKSLSPMPQIAGVTYYGDIPKQPKRVVSLA 64

15 Query: 62 STYTGYLKKLDMNLVGVTSYDKKNPILAKTVKKAKQVAATDLEAVTTLKPDLLIVVGSTEE 121
 +TG + LD+N+VGV ++ K NP + ++ +V+ +LE + L PDLI+ ST +
 Sbjct: 65 --FTGNILALDVNVVGVDTSKNNPNYEQLLQDVTEVSEENLEQIMELDPDLIIAYSTVQ 122

20 Query: 122 NIKQLAEIAPVISIEYRKRQVLSDFGRIFNKEDKAKKWLKDWKTCTAAYEKEVKAVT 181
 N +QL EIAP + Y DYLL + G++ NKE++A+ W+ D+K + +E+K
 Sbjct: 123 NAEQLQEIAPTPLYTYNNLDYLEQHVIEIGKLLNKEEEAQAWVDDFKARAEQAGEEIKEKI 182

25 Query: 182 GDKATFTIMGLYEKDVLYFGKDWGRGGEIIHQAFHYDAPEKVKTEVFKQGYLSLSQEVLP 241
 G+ AT +++ +E +Y+FG +WGRG EI++Q PE+V+ GY +LS E LP
 Sbjct: 183 GEDATVSVIETFDQLYVFGNNGRGTILYQIMDLAMPERVEEMALADGYALSFEALP 242

30 Query: 242 DYIGDYVVAAEDDKTGSALYESKLWQSIPAVKKHHVIKVNANVFYFTDPLSLEYQLETL 301
 ++ GDY+++ +++D+ ++ E+ +QSIPAV+ V + NA FYF DPLSLE QLE
 Sbjct: 243 EFAGDYIIL-SKNDEADNSFQETNTYQSIPAVQNGQVFANAKEFYFNDPLSLELQLEFF 301

Query: 302 REAILS 307
 +E LS
 Sbjct: 302 KEHFLS 307

An alignment of the GAS and GBS proteins is shown below.

Identities = 140/316 (44%), Positives = 212/316 (66%), Gaps = 12/316 (3%)

35 Query: 1 MKKIGIIV-LTLLTFFLVSCGQQTQESTKTT--ISKMPKIEGFTYYGKIPENPKKVINF 57
 MKK+ +++ L L T L++CG Q S + +S MP+I G TTYG IP+ PK+V++
 Sbjct: 1 MKKLTLTLTCLTTITLIACGNQATNHSNTASKSLSPMPQIAGVTYYGDIPKQPKRVVSL 60

40 Query: 58 TYSYTGILLKLGVN---VSSYSLDLEKDSFVFGKQLKEAKKLTADDTEAIAAQKPDLMIV 114
 +YTGYL KL +N V+SY +K +P+ K +K+AK++ A D EA+ KPDLI+V
 Sbjct: 61 ASTYTGYLKKLDMNLVGVTSY---DKKNPILAKTVKKAKQVAATDLEAVTTLKPDLLIVV 116

45 Query: 115 FDQDPNINTLKKIAPTLVIKYGAQNYLDMMPALGKVFGEKEANQWVSQWTKTTLAVK 174
 + NI L +IAP + I+Y ++YL ++ G++F KE +A +W+ WTKT A +K+
 Sbjct: 117 GSTEENIKQLAEIAPVISIEYRKRQVLSDFGRIFNKEDKAKKWLKDWKTCTAAYEKE 176

50 Query: 175 LHHILKPNTTFTIMDFYDKNIYLYGNNGRGGELIYDSLGYAAPEKVKDVFKKGWFTVS 234
 + + TFTIM Y+K++YL+G ++GRGGE+I+ + Y APEKVK +VFK+G+ ++S
 Sbjct: 177 VKAVTGDKATFTIMGLYEKDVLYFGKDWGRGGEIIHQAFHYDAPEKVKTEVFKQGYLSLS 236

55 Query: 235 QEAGIDYVGDYALVNINKTTKKAASSLKESDVWKNLPAVKKGHIIESNYDVYFSDPLSL 294
 QE + DY+GDY +V K S+L ES +W+++PAVKK H+I+ N +VFYF+DPLSL
 Sbjct: 237 QEVLDPYIGDYVVAAE--DDKTGSALYESKLWQSIPAVKKHHVIKVNANVFYFTDPLSL 294

Query: 295 EAQLKSFTKAIKENTN 310
 E QL++ +AI + N
 Sbjct: 295 EYQLETLREALSEN 310

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

-1333-

Example 1187

A DNA sequence (GBSx1263) was identified in *S.agalactiae* <SEQ ID 3695> which encodes the amino acid sequence <SEQ ID 3696>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1188

A DNA sequence (GBSx1264) was identified in *S.agalactiae* <SEQ ID 3697> which encodes the amino acid sequence <SEQ ID 3698>. This protein is predicted to be ferrichrome transport permease (permease). Analysis of this protein sequence reveals the following:

```

20 Possible site: 39
>>> May be a lipoprotein
      INTEGRAL    Likelihood = -12.74    Transmembrane    129 - 145 ( 123 - 150)
      INTEGRAL    Likelihood = -10.67    Transmembrane    248 - 264 ( 240 - 283)
      INTEGRAL    Likelihood = -10.14    Transmembrane    205 - 221 ( 196 - 228)
25      INTEGRAL    Likelihood = -5.95    Transmembrane    319 - 335 ( 317 - 336)
      INTEGRAL    Likelihood = -3.56    Transmembrane    73 - 89 ( 73 - 90)
      INTEGRAL    Likelihood = -3.19    Transmembrane    288 - 304 ( 288 - 304)
      INTEGRAL    Likelihood = -2.76    Transmembrane    266 - 282 ( 265 - 283)
      INTEGRAL    Likelihood = -2.23    Transmembrane    103 - 119 ( 101 - 122)
30      INTEGRAL    Likelihood = -1.01    Transmembrane    158 - 174 ( 158 - 174)

----- Final Results -----
      bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF98154 GB:AF251216 FhuB [Staphylococcus aureus]
Identities = 116/313 (37%), Positives = 194/313 (61%), Gaps = 3/313 (0%)
40
Query: 26 ILFLIGCYASLRFGAINFKTSDLITVLKNPLKNSNAQDVIFDIRLPRIIAAILVGAAMSQ 85
      ++ LI + S G + S + I + N ++ Q++I +IR+PR IAA++VG A++
Sbjct: 28 MILLITLFI STLIGDAKI QASTIIEAIFNYPNPSNQQNIINEIRIPRNIAAVIVGMALAV 87

45
Query: 86 AGAIMQGVTRNAIADPGLGINAGAGLALVVAYAFGLSMHYSTILIVCLLGSVISCLLVF 145
      +GAI+QGVTRN +ADP L+G+N+GA AL + YA L + + ++ LG+++ +V
Sbjct: 88 SGAI IQGVTRNGLADPALIGLNSGASFALALTYAVLPNTSFLILMFAGFLGAILGGAIVL 147

50
Query: 146 TLSYTKQKGYHQLRLILAGAMISTLFTSVGQVVTLYFKLNRTVIGWQAGGLSQINWKMLI 205
      + +++ G++ +R+ILAGA +S + T++ Q + L F+LN+TV W AGG+S W L
Sbjct: 148 MIGRSRRDGFNPMRIILAGA AVSAML TALSQGI ALAFRLNQT VTFWTAGGVSGTTWSHLK 207

Query: 206 IIAPIIILGLLISQLLAHQLTILSLNESVAKALGQKTQLMTAFLLLLIVLFLSASSVALIG 265
      P+I + L I ++ QLTI L+L ES+AK LGQ ++ L+I + L+ +VA+ G
55 Sbjct: 208 WAIP LIGIALFIILTISKQLTILNLGESLAKGLGQNV TIRGICLI IAMILAGI AIAIAG 267

```

-1334-

Query: 266 TVSFGLIIPHFIKLFIPKDYRLLLPLIGFSGATFMIWVDLSSRIINPPSETSISSIISI 325
 V+F+GL++PH + I DY +LPL G ++ D+ +R + E + +IIS
 Sbjct: 268 QVAFVGLMVPFIARFLIGTDYAKILPLTALLGGILVLVADVIARYL---GEAPVGAIISF 324

5 Query: 326 VGLPCFLWLIRKG 338
 +G+P FL+L++KG
 Sbjct: 325 IGVFPYFLYLKKG 337

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3699> which encodes the amino acid
 10 sequence <SEQ ID 3700>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.09	Transmembrane	256 - 272 (248 - 287)
INTEGRAL	Likelihood = -10.67	Transmembrane	26 - 42 (23 - 48)
INTEGRAL	Likelihood = -6.90	Transmembrane	137 - 153 (133 - 157)
INTEGRAL	Likelihood = -5.10	Transmembrane	167 - 183 (166 - 187)
INTEGRAL	Likelihood = -4.57	Transmembrane	213 - 229 (210 - 232)
INTEGRAL	Likelihood = -2.02	Transmembrane	112 - 128 (110 - 131)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:AAF98154 GB:AF251216 FhuB [Staphylococcus aureus]
 Identities = 99/274 (36%), Positives = 159/274 (57%), Gaps = 1/274 (0%)

30 Query: 34 LSFSLCVAIYCHLRFGAVALSHQDLNSILFG-KQNGHKANVLLAIRLPRLFGATLTGSAL 92
 LS L + ++ G + + +F + + N++ IR+PR A + G AL
 Sbjct: 26 LSMILLITLFIISTLIGDAKIQASTIIEAIFNYPNSNQQNIINEIRIPRNIAAVIVGMAL 85

35 Query: 93 AVSGTIMQAITRNPPIAEPGLLGINAGAGLALVLAYAFVPHLHYSLLIILLSSSLAATL 152
 AVSG I+Q +TRN +A+P L+G+N+GA AL L YA +P+ + +++ LG+ L +
 Sbjct: 86 AVSGAIIQGVTRNGLADPALIGLNSGASFALALTYAVLPNTSFLILMFAGFLGAILGGAI 145

40 Query: 153 VFGLSYQSGKGYHQLRLVLGAMVSILLSALGQGITNYYHLANAVIGWQAGGLVGVNWQM 212
 V + G++ +R++LAGA VS +L+AL QGI + L V W AGG+ G W
 Sbjct: 146 VLMIGRSRRDGFNPMRIILAGAAVSAMLTALSQGIALLAFRLNQTFTFWTAGGVSGTWSH 205

45 Query: 213 IGYIAPLIILSLCLAQLLSYHLTVLSLSESQAKALGQKTNLISAVFMILVLILSSAAVAI 272
 + + PLI ++L + +S LT+L+L ES AK LGQ +I + +I+ +IL+ AVAI
 Sbjct: 206 LKWAIPILGIALFIILTISKQLTILNLGESLAKGLQNVMTIRGICLIIAMILAGIAVAI 265

Query: 273 AGSISFIGLVIPHLMKHFTPHHYRYLLPLCAVSG 306
 AG ++F+GL++PH+ + Y +LPL A+ G
 Sbjct: 266 AQVAFVGLMVPFIARFLIGTDYAKILPLTALLG 299

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 158/295 (53%), Positives = 214/295 (71%), Gaps = 1/295 (0%)

Query: 6 KKL VQKNKSNHFWLVFFITLILFLIGCYASLRFGAINFKTSDLITVLKNPLKNSNAQDVI 65
 KK KS+ FWLVF + + Y LRFGA+ DL ++L +N + +V+
 Sbjct: 16 KKTQIITKSHIFWLVFVLLSFSLCVAIYCHLRFGAVALSHQDLNSILFGK-QNGHKANVL 74

55 Query: 66 FDIRLPRIIAAILVGAAMSQAGAIMQGVTRNAIADPGLLGINAGAGLALVVAYAFGLSMH 125
 IRLPR+ A L G+A++ +G IMQ +TRN IA+PGLLGINAGAGLALV+AYAF+ +H
 Sbjct: 75 LAIRLPRLFGATLTGSALAVSGTIMQAITRNPPIAEPGLLGINAGAGLALVLAYAFVPHLH 134

60 Query: 126 YSTILIVCLLGSVISCLLVFTLSYTKQKGYHQLRLILAGAMISTLFTSVGQVVTLYFKLN 185
 YS I+++ LLGS ++ LVF LSY KGYHQLRL+LAGAM+S L +++GQ +T Y+ L
 Sbjct: 135 YSLIILLSSLLGSSLAATLVFGLSYQSGKGYHQLRLVLGAMVSILLSALGQGITNYYHLA 194

Query: 186 RTVIGWQAGGLSQINWKMLIIITAPIIILGLLISQLLAHQTLTILSLNESVAKALGQKTQLM 245

-1335-

VIGWQAGGL +NW+M+ IAP+IIL L ++QLL++ LT+LSL+ES AKALGQKT L+
 Sbjct: 195 NAVIGWQAGGLVGVNWQMIGYIAPLIILSLCLAQLLSYHLTVLSLSESQAKALGQKTNLI 254

Query: 246 TAFLLLLIVLFLSASSVALIGTVSFIGLIIPHFIKLFIPKDYRLLLPLIGFSGATF 300
 +A +++VL LS+++VA+ G++SFIGL+IPH +K F P YR LLPL SGA+F
 Sbjct: 255 SAVFMILVLILSSAAVAIAGSISFIGLVIPHLMKHFTPHHYRYLLPLCAVSGASF 309

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1189

A DNA sequence (GBSx1265) was identified in *S.agalactiae* <SEQ ID 3701> which encodes the amino acid sequence <SEQ ID 3702>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1492(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 1190

A DNA sequence (GBSx1266) was identified in *S.agalactiae* <SEQ ID 3703> which encodes the amino acid sequence <SEQ ID 3704>. This protein is predicted to be ferrichrome transport permease (permease). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -10.35	Transmembrane	282 - 298 (279 - 309)
INTEGRAL	Likelihood = -7.06	Transmembrane	120 - 136 (115 - 141)
INTEGRAL	Likelihood = -7.01	Transmembrane	62 - 78 (61 - 80)
INTEGRAL	Likelihood = -6.10	Transmembrane	250 - 266 (241 - 272)
INTEGRAL	Likelihood = -5.52	Transmembrane	196 - 212 (190 - 215)
INTEGRAL	Likelihood = -5.47	Transmembrane	155 - 171 (151 - 174)
INTEGRAL	Likelihood = -4.99	Transmembrane	304 - 320 (303 - 322)
INTEGRAL	Likelihood = -3.35	Transmembrane	91 - 107 (90 - 110)

----- Final Results -----
 bacterial membrane --- Certainty=0.5140(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98155 GB:AF251216 FhuG [Staphylococcus aureus]
 Identities = 122/334 (36%), Positives = 208/334 (61%), Gaps = 3/334 (0%)

Query: 1 MIQKNKAPFVLISSVIILLLLILV---SISLGYANTSVIDVLKLSIGKSDDAFLFIITNI 57
 MI N LI+ + +LL L SI+ G N V K + G+ D I+ +
 Sbjct: 1 MISSNNKRRLIALAVFSILLFLGCTWSITSGEYNIPVERFFKTLIGQDAIDELILLDF 60

Query: 58 RLPRIIVCIFGGASLGIAGLLQLTLTKNPLADSGILGINAGAGLVIALTIGTFNVSNPIT 117
 RLPR+++ I GA+L I+G ++Q++TKNP+A+ GILGINAG G IAL I ++

-1336-

Sbjct: 61 RLPRMMITILAGAALSISGAIVQSVTKNPIAEPGILGINAGGGFAIALFIAIGKINADNF 120

Query: 118 LYFLPLFAMFGGLVTIFLIYLSYRRNHNISPTRLIVTGIGISTIISGVMILIISQSNNQ 177
 +Y LPL ++ GG+ T +I++ S+ +N ++P +++ G+G+ T + G I I+S+ +++
 Sbjct: 121 VVVLPLISILGGITTALIIIFISFNKNEGVTPASMLIGVGLQTALYGGSTITMSKFDK 180

Query: 178 KMDMIVEWLSGKITISSWTTITTFIPILILLWGLAYSRSRHLNIMNLNEQTALALGLHLK 237
 + D I W +G I W +I F+P ++++ +S LNI++ + A LG+ L
 Sbjct: 181 QSDFIAAWFAGNIWGEWPFVIAFLPWVLIIPYLLPKSNTLNIIHTGDNIARGLGVRLS 240

Query: 238 KERIYTLMLTSSLAISVVLIGNITFIGLLAGHLRRLGNHKKIILPSCLLIGAILLV 297
 +ER+ + L++ +V + G+I+FIGL+ H+++R++G H++ LP +L+GA +L++
 Sbjct: 241 RERLILFFIAVMLSSAAVAVAGSISFIGLMGPHIAKRIVGPRHQLFLPIAILVGACLLVI 300

Query: 298 SDTIGRLLLVGTGIPTGLVVSIIIGAPYFLWLMTK 331
 +DTIG+++L G+P G+VV+IIGAPYFL+LM K
 Sbjct: 301 ADTIGKIVLPQGGVPAGIVVAIIGAPYFLYLYMYK 334

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1939> which encodes the amino acid sequence <SEQ ID 1940>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.93	Transmembrane	254 - 270 (252 - 284)
INTEGRAL	Likelihood = -10.46	Transmembrane	294 - 310 (292 - 320)
INTEGRAL	Likelihood = -6.74	Transmembrane	25 - 41 (18 - 43)
INTEGRAL	Likelihood = -6.26	Transmembrane	103 - 119 (102 - 125)
INTEGRAL	Likelihood = -3.66	Transmembrane	164 - 180 (164 - 186)
INTEGRAL	Likelihood = -3.03	Transmembrane	209 - 225 (207 - 226)
INTEGRAL	Likelihood = -2.71	Transmembrane	74 - 90 (74 - 91)
INTEGRAL	Likelihood = -2.13	Transmembrane	326 - 342 (325 - 343)
INTEGRAL	Likelihood = -1.97	Transmembrane	135 - 151 (135 - 151)

----- Final Results -----

bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/322 (47%), Positives = 229/322 (70%), Gaps = 1/322 (0%)

Query: 11 LISSVIIILLLLIL-VSISLGYANTSVIDVLKLISGKSDDAFLFIITNIRLPRIIVCIFGG 69
 L +S+I+LL+ ++ +++SLG ++ S +D++ + GKS A FI+ NIRLPRI+ GG
 Sbjct: 22 LYTSLILLVSLMGLALSIGESHLSFLDLVHVFLGKSSHAISFIVINIRLPRIILAACLG 81

Query: 70 ASLGIAGLLLQTLTKNPLADSGILGINAGAGLVIALTIGTFNVSNTILYFLPLFAMFGG 129
 SL ++GLLLQ LT+NPLADSG+LGI GAG+ +A+ + I ++LPLFAM G
 Sbjct: 82 GSLALSGLLLQRLTRNPLADSGVLGITIGAGISLAIVVSFSFFEQAHSYHLPLFAMLGA 141

Query: 130 LVTIFLIYLSYRRNHNISPTRLIVTGIGISTIISGVMILIISQSNNQKMDMIVEWLSGK 189
 +VT F +Y +S + I PTRLI+TG+ ++T++S +M+ ++ N K+D+++ WLSG+
 Sbjct: 142 IVTTFSVYWLSTKQGQIDPTRLILTGVAVTMLSSLMVALVGHINRYKVDLVINWLSGQ 201

Query: 190 ITISSWTTITTFIPILILLWGLAYSRSRHLNIMNLNEQTALALGLHLKKERIYTLMLTSS 249
 + W T+ P+L+ W L YS++ LNIM L + TA+ LGL L ++R L+L +
 Sbjct: 202 LIGDDWPTLSVIAPLLLCFWLLTYSQAHLNIMGLADNTAIGLGLPLNRKRRLILVLAAG 261

Query: 250 LAASVVLIGNITFIGLLAGHLRRLGNHKKIILPSCLLIGAILLVSDTIGRLLLVGT 309
 L A+SV+L+GNI+FIGL+AGH S L+G+NHKI +P +LIG I+LLV+DT+GR+ LVG+
 Sbjct: 262 LGALSVLLVGNISFIGLIAGHFSTYLVGSNHKITIPISILIGMILLLVADTVGRVYLVGS 321

Query: 310 GIPTGLVVSIIIGAPYFLWLMTK 331
 I TG++VS+IGAPYFL+LM K
 Sbjct: 322 NIQTGILVSLIGAPYFLYLYMAK 343

There is also homology to SEQ ID 396.

-1337-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1191

A DNA sequence (GBSx1267) was identified in *S.agalactiae* <SEQ ID 3705> which encodes the amino acid sequence <SEQ ID 3706>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3785(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC05779 GB:AF051356 unknown [Streptococcus mutans]

Identities = 49/93 (52%), Positives = 63/93 (67%)

Query: 1 MILTFNPGKLERQEFFKELINYLWIHDDVTLRKIKSHFTDYSKIDRLLEEYINHG YILRQ 60

MI +N KL RQ FF +LINYL IHDDVTLR+IK +F D ++R +E+Y+ GY+LR+

Sbjct: 1 MIKIYNGDKLTRQPPFIKLINYLQIHDDVTLRQIKRNFADTEHLERSIEDYVQAGYVLRE 60

Query: 61 NKRYSLNLPFLSSLDGLVLDLDFIDSDSQIYQ 93

NK Y L +LDGL LD +F+D S IYQ

Sbjct: 61 NKHYNAFELLENLDGLTLDSDQIFVDDQSSIYQ 93

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3707> which encodes the amino acid sequence <SEQ ID 3708>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3447(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 108/212 (50%), Positives = 143/212 (66%)

Query: 1 MILTFNPGKLERQEFFKELINYLWIHDDVTLRKIKSHFTDYSKIDRLLEEYINHG YILRQ 60

MI F+ KL RQ FF++LINYL HD V LR+IK F + + ID+ +E Y+ GYI R+

Sbjct: 1 MITVPHSDKLTRQPPFQDLINYLQHDHVLREIKKAFPNVTGIDKAIESYVQAGYIRRE 60

Query: 61 NKRYSLNLPFLSSLDGLVLDLDFIDSDSQIYQLLQKRKFVTNLDNPTNHLVFEETDFE 120

NKRY +NLP +SS L LD ++F+D+ S +Y+ + F T L N TN ++ E+T+

Sbjct: 61 NKRYGINLPLVSSDQQLALDTMLFVDTCSAMYENILAVVFETQLTNQTNRVMIKEKINIT 120

Query: 121 RNTLTLSNIFYKLTNGYPLSREQKKLYQLLGDVNSEYALKYMSSFILKFLRKDSVKQKRT 180

R+ L/L+NIFY+L G S EQ LY LLGDVN EYALKYM++F+LKF RKD V QKR

Sbjct: 121 RDDLTLANIFYRLKRGEKPSAEQMDLYDLLGDVNQEYALKYMTTFLKFKTRKDFVMQKRP 180

Query: 181 VIFIQALELLGYISLNQDTTYRLNAKL DVEAL 212

IF++AL LGY+ + TTY+L LD E+L

Sbjct: 181 DIFVEALVTGLGYLKQVEPTTYQLLMTLDKESL 212

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1338-

Example 1192

A DNA sequence (GBSx1268) was identified in *S.agalactiae* <SEQ ID 3709> which encodes the amino acid sequence <SEQ ID 3710>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 24
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0824(Affirmative) < succ>
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:AAB39104 GB:U57759 intrageneric coaggregation-relevant adhesin
      [Streptococcus gordonii]
15      Identities = 261/311 (83%), Positives = 283/311 (90%)

      Query: 1  MSKILVFGHQNPDSDAIGSSVAFAYLAKEAWGLDTEAVALGTPNEETAYVLDYFGVQAPR 60
      MSKILVFGHQNPDSDAIGSS AFAYLA+EA+GLDTEAVALG PNEETA+VLDYFGV APR
      Sbjct: 1  MSKILVFGHQNPDSDAIGSSYAFAYLAREAYGLDTEAVALGEPNEETA+VLDYFGVAAAPR 60

20      Query: 61  VVESAKAEGVETVILTDHNEFQQSISDIKDVTYGVVDHHRVANFETANPLYMRLEPVGS 120
      V+ SAKAEG E VILTDHNEFQQS++DI +V VYGVVDHHRVANFETANPLYMRLEPVGS
      Sbjct: 61  VITSAKAEGAEQVILTDHNEFQQSVADIAEVEVYGVVDHHRVANFETANPLYMRLEPVGS 120

25      Query: 121 ASSIVYRMFKENGVSVPKELAGLLSGLISDTLLLLKSPTHASDIPVAKELAELAGVNLE 180
      ASSIVYRMFK+ V+V KE+AGL+LSGLISDTLLLLKSPTH +D +A ELAELAGVNLE
      Sbjct: 121 ASSIVYRMFKEHSAVSVKEIAGLMLSGGLISDTLLLLKSPTHPTDKAIAPELAELAGVNLE 180

30      Query: 181 EYGLEMLKAGTNLSSKTAELIDIDAKTFELNGEAVRVAQVNTVDINDILARQEEIEVAI 240
      EYGL MLKAGTNL+SK+A ELIDIDAKTFELNG VRVAQVNTVDI ++L RQ EIE AI
      Sbjct: 181 EYGLAMLKAGTNLASKSAEELIDIDAKTFELNGNNVRVAQVNTVDIAEVLERQAEIEAAI 240

      Query: 241 QEAIVTEGYSDFLMITDIVNSNSEILALGSNMAKVEAAFEFTLENNHAFLAGAVSRKKQ 300
      ++AI GYSDFLMITDI+NSNSEILA+GSNM KVEAAF F LENNHAFLAGAVSRKKQ
35      Sbjct: 241 EKAIADNGYSDFLMITDIINSNSEILAIGSNMDKVEAAFNFLVLENNHAFLAGAVSRKKQ 300

      Query: 301 VVPQLTESYNA 311
      VVPQLTES+NA
      Sbjct: 301 VVPQLTESFNA 311
40

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3711> which encodes the amino acid sequence <SEQ ID 3712>. Analysis of this protein sequence reveals the following:

```

      Possible site: 52
      >>> Seems to have no N-terminal signal sequence
45      INTEGRAL    Likelihood = -2.02    Transmembrane  141 - 157 ( 141 - 157)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9103> which encodes the amino acid sequence <SEQ ID 9104>. Analysis of this protein sequence reveals the following:

```

      Possible site: 50
55      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.02    Transmembrane  139 - 155 ( 139 - 155)

      ----- Final Results -----
      bacterial membrane --- Certainty= 0.181(Affirmative) < succ>
60      bacterial outside --- Certainty= 0.000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

```


-1339-

An alignment of the GAS and GBS proteins is shown below.

Identities = 253/311 (81%), Positives = 283/311 (90%)

```

5  Query: 1  MSKILVFGHQNPDSDAIGSSVAFAYLAKEAWGLDTEAVALGTPNEETAYVLDYFGVQAPR 60
      MSKILVFGHQNPDAI SS AF YL+++A+GLDTE VALGTPNEETA+ LDYFGV+APR
      Sbjct: 3  MSKILVFGHQNPDTDAIASSYAFDYLQKAFGLDTEVVALGTPNEETAFAFDYFGVEAPR 62

10 Query: 61  VVESAKAEGVETVILTDHNEFQQSISDIKDVTVYGVVDHHRVANFETANPLYMRLEPVGS 120
      VVESAKA+G E VILTDHNEFQQSI+DI++V VYGVVDHHRVANFETANPLYMR+EPVGS
      Sbjct: 63  VVESAKAQGSEQVILTDHNEFQQSIADIREVEVYGVVDHHRVANFETANPLYMRVEPVGS 122

15 Query: 121  ASSIVYRMFKENGVSVPKELAGLLSGLISDTLLKSPTHASDIPVAKELAEAGVNLE 180
      ASSIVYRMFKENG+ VPK +AG+LLSGLISDTLLKSPTH SD VA+ELAE LA VNLE
      Sbjct: 123  ASSIVYRMFKENGIEVPKAIAGMLLSGLISDTLLKSPTHVSDHLVAEELAE LAE VNLE 182

20 Query: 181  EYGLEMLKAGTNLSSKTAELIDIDAKTFELNGEAVRVAQVNTVDINDILARQEEIEVAI 240
      +YG+ +LKAGTNL+SK+ ELI IDAKTFELNG AVRVAQVNTVDI ++L RQE IE AI
      Sbjct: 183  DYGMALLKAGTNLASKSEVELIGIDAKTFELNGNAVRVAQVNTVDIAEVLERQEAIEAAI 242

25 Query: 241  QEAIIVTEGYSDFVLMITDIVNSNSEILALGSNMAKVEAAFEFTLENNHAFLAGAVSRKKQ 300
      ++A+ EGYSDFVLMITDIVNSNSEILA+G+NM KVEAAF FTL+NNHAFLAGAVSRKKQ
      Sbjct: 243  KDAMAAEGYSDFVLMITDIVNSNSEILAIGANMDKVEAAFNFTLDNNHAFLAGAVSRKKQ 302

30 Query: 301  VVPQLTESYNA 311
      VVPQLTES+ A
      Sbjct: 303  VVPQLTESFGA 313

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1193

A DNA sequence (GBSx1269) was identified in *S.galactiae* <SEQ ID 3713> which encodes the amino acid sequence <SEQ ID 3714>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2769(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC05773 GB:AF051356 pyruvate-formate lyase activating enzyme
[Streptococcus mutans]

Identities = 184/260 (70%), Positives = 217/260 (82%)

```

45 Query: 3  EIDYKKVTGMIHSTESFGSVDGPGIRFIIFMQGCKMRCQYCHNPDTWEMETNNSKERTVE 62
      ++DY+KVTG+++STESFGSVDGPGIRF++FMQGC+MRCQYCHNPDTW M+ + + ERT
      Sbjct: 4  KVDYKVTGLVNSTESFGSVDGPGIRFVVFVMQGCQMRQYCHNPDTWAMKNDRATERTAG 63

50 Query: 63  DVLKEALRYKHFWDGKITVSGGEAMLQIDFITALFIEAKKLGIIHTTLDTCGFAYRATP 122
      DV KEALR+K FWG GGITVSGGEA LQ+DF+ ALF AK+ GIHTTLDTC +R TP
      Sbjct: 64  DVFKEALRFKDFWGDITVSGGEATLQMDFLIALFSLAKEKGIHTTLDTCALTFRNTPT 123

55 Query: 123  EYHAILEKLLDVTDLVLLDLKEIDSEQHKIVTRQSNKNILQFARYLSDRGTPVWIRHVLV 182
      +Y EKL+ VTDLVLLD+KEI+ +QHKIVT SNK IL ARYLS D G PVWIRHVLV
      Sbjct: 124  KYLEKYEKLMVTDLVLLDIKEINPDQHKIVTGHSNKTILACARYLSDIGKPVWIRHVLV 183

60 Query: 183  PGLTDDIDHLKRLGEFVQTLDNVDKFEVLPYHTMGEFKWRELGIPIYPLAGVKPPTPERVK 242
      PGLTD D+ L +LGE+V+TL NV +FE+LPYHTMGEFKWRELGIPIYPL GVKPPTP+RV+
      Sbjct: 184  PGLTDRDEDLIKGEYVKTLLKNVQRFELPYHTMGEFKWRELGIPIYPLEGVKPPPTDPRVR 243

```

-1340-

Query: 243 NAKDIMKTESYTEYLKRIQN 262
 NAK +M TE+Y EY KRI +
 Sbjct: 244 NAKKLMHTETYEYKKRINH 263

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3715> which encodes the amino acid sequence <SEQ ID 3716>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4614(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 223/260 (85%), Positives = 239/260 (91%)

Query: 1 MAEIDYKKVTGMIHSTESFGSDGPGIRFIIFMQGCKMRCQYCHNPDTWEMETNNSKERT 60
 M E DY +VTGM+HSTESFGSDGPGIRFIIF+QGCK+RCQYCHNPDTWEMETNNSK RT
 Sbjct: 25 MTEKDYGQVTGMVHSTESFGSDGPGIRFIIFLQGCKLRCQYCHNPDTWEMETNNSKIRT 84
 Query: 61 VEDVLKEALRYKHFVGKDGGITVSGGEAMLQIDFITALFIEAKKLGIIHTTLDTCGFAYRA 120
 V DVLKEAL+YKHFVGK GGITVSGGEAMLQIDFITALFIEAKKLGIIHTTLDTCGF YR
 Sbjct: 85 VNDVLKEALQYKHFVGKGGGITVSGGEAMLQIDFITALFIEAKKLGIIHTTLDTCGFTYRP 144
 Query: 121 TPEYHAILEKLLDVTDLVLLDLKEIDSEQHKIVTRQSNKNILQFARYLSDRGTFVWIRHV 180
 TPEYH +L+ LL VTDL+LLDLKEID +QHKIVTRQ NKNILQFARYLSD+ PVWIRHV
 Sbjct: 145 TPEYHQVLDNLLAVTDLILLDLKEIDKQHKIVTRQPNKNILQFARYLSDKQIPVWIRHV 204
 Query: 181 LVPGLTDIDDHLKRLGEFVQTLDNVDKFEVLPYHTMGFEFKWRELGIPYPLAGVKPPTPER 240
 LVPGLTDIDDHL RLGEFV+TL NVDKFEVLPYHTMGFEFKWRELGIPY L GVKPPT ER
 Sbjct: 205 LVPGLTDIDDHLTRLGEFVKTLKNVDKFEVLPYHTMGFEFKWRELGIPYQLEGVKPPTKER 264
 Query: 241 VKNAKDIMKTESYTEYLKRI 260
 V+NAK+M+TESYTEY+ RI
 Sbjct: 265 VQNAKNLMQTESYTEYMNRI 284

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1194

A DNA sequence (GBSx1270) was identified in *S.agalactiae* <SEQ ID 3717> which encodes the amino acid sequence <SEQ ID 3718>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -7.06 Transmembrane 105 - 121 (103 - 126)
 INTEGRAL Likelihood = -5.57 Transmembrane 137 - 153 (136 - 162)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC05772 GB:AF051356 putative hemolysin [Streptococcus mutans]
 Identities = 347/445 (77%), Positives = 406/445 (90%), Gaps = 1/445 (0%)
 Query: 1 MQDPGSQSLLLQFVILLILTLFNAFFSASEMALVSLNRSKVEQKAEEGDKRYRRLLDVLE 60
 M+DPGSQSL+LQF++LLILTL NAFSA+EMALVSLNR++VEQKAEEG+K+Y RLL VLE

-1341-

Sbjct: 1 MEDPGSQSLILQFLLLLILTLTLCNAFFSATEMALVSLNRRARVEQKAEEGEKKYIRLLKVL 60

Query: 61 NPNNFLSTIQVGITFISLLQGSLSASLGHVISGWLGN SATARTAGSIIALIFLTYVSIV 120
NPNNFLSTIQVGIT I+LL GASL+ SLG I+ W GNSATARTAGS+I+L FLTY+SIV

5 Sbjct: 61 NPNNFLSTIQVGITLITLLSGASLADSLGREIAVWFGNSATARTAGSLISLAFLTYISIV 120

Query: 121 LGELYPKRIAMNLKDRIAIVSAPIIIFLGKIVSPFVWLLSASTNLLSRITPMTFDDADEK 180
LGELYPKRIAMNLK+ LA++SAP+IIFLGK+VSPFVWLLS STNLLSR+TPMTFDDADEK

10 Sbjct: 121 LGELYPKRIAMNLKENLAVLSAPVIFLGKIVSPFVWLLSVSTNLLSRITPMTFDDADEK 180

Query: 181 MTRDEIEYMLTNSEETLEAEIEMLQGFSLDEMAREVMVPRTD AFMIDINND AQSNIE 240
MTRDEIEYMLTNSEETL+A+EIEMLQGFSLDE+MAREVMVPRTD AFM+DIN+D+ I+

Sbjct: 181 MTRDEIEYMLTNSEETLDADEIEMLQGVFSLDELMAREVMVPRTD AFMVDINDDSSDIIQ 240

15 Query: 241 GILSQNF SRVPVFD DDKDRVVGVLHTKRLLEAGFKTGFD TIDLKILQEPLFVPETIFVD 300
IL++ FSR+PV+DDDKD+++G++HTK LL AGFK GFD I+LR+ILQEPLFVPETI V+

Sbjct: 241 TILNERFSRIPVYDDDKDKIIGIHTKNLLNAGFKEGFDHINLRRILQEPLFVPETIVVN 300

Query: 301 DLLKALRNTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDETDTAEQFVREIDENIYI 360
DLL AL+NTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDET VREI +N YI

20 Sbjct: 301 DLLTALKNTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDETDKTAISVREIADNTYI 360

Query: 361 VLGTMTLNEFNDFETELSDVDVTIAGYYLTGVGSIPNQEEKVAYEVD SKDKHITLIND 420
VLGTMTLN+FN+YFET+LES D+VDTIAG+YLTGVG+IP+QEEK +EV+S KH+ LIND

25 Sbjct: 361 VLGTMTLNDNFENYFETDLESDNVDVTIAGFYLTGVGTIP SQEKEHFEVESNGKHLELIND 420

Query: 421 KVKDGRITKLKVLVLSDIEQ-NIEKD 444
KVKDGR+TKLK+L+S++E+ EKD

30 Sbjct: 421 KVKDGRVTKLKILVSEVEEKEDEKD 445

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3719> which encodes the amino acid sequence <SEQ ID 3720>. Analysis of this protein sequence reveals the following:

Possible site: 42

35 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.76	Transmembrane	22 - 38 (16 - 47)
INTEGRAL	Likelihood = -5.57	Transmembrane	118 - 134 (117 - 138)
INTEGRAL	Likelihood = -3.19	Transmembrane	150 - 166 (149 - 169)

40 ----- Final Results -----

bacterial membrane	---	Certainty=0.4503(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

45 The protein has homology with the following sequences in the databases:

>GP:AAC05772 GB:AF051356 putative hemolysin [Streptococcus mutans]
Identities = 343/443 (77%), Positives = 401/443 (90%)

50 Query: 14 MEDPVSQSLVIQFLLLVLTLLNAFFSASEMALVSLNRSRVEQKAADGDKKYARLLRVLE 73
MEDP SQSL++QFLL++LTL NAFFSA+EMALVSLNR+RVEQKA +G+KKY RLL+VLE

Sbjct: 1 MEDPGSQSLILQFLLLLILTLTLCNAFFSATEMALVSLNRRARVEQKAEEGEKKYIRLLKVL 60

Query: 74 EPNHFLSTIQVGITFISLLSGASLSASLGKVISGWLGN SATARTAGTIISLVFLTYVSIV 133
PN+FLSTIQVGIT I+LLSGASL+ SLG+ I+ W GNSATARTAG++ISL FLTY+SIV

55 Sbjct: 61 NPNNFLSTIQVGITLITLLSGASLADSLGREIAVWFGNSATARTAGSLISLAFLTYISIV 120

Query: 134 LGELYPKRIAMNLKDKLAIVSAPIIIGLGRVSPFVWLLSASTNLLSRITPMTFDDADEQ 193
LGELYPKRIAMNLK+ LA++SAP+II LG++VSPFVWLLS STNLLSRITPMTFDDADE+

60 Sbjct: 121 LGELYPKRIAMNLKENLAVLSAPVIFLGKIVSPFVWLLSVSTNLLSRITPMTFDDADEK 180

Query: 194 MTRDEIEYMLSKSEATLDAEIEMLQGVFSLDEMAREVMVPRTD AFMIDINDDPLENIQ 253
MTRDEIEYML+ SE TLDA+EIEMLQGVFSLDE+MAREVMVPRTD AFM+DINDD + IQ

Sbjct: 181 MTRDEIEYMLTNSEETLDADEIEMLQGVFSLDELMAREVMVPRTD AFMVDINDDSSDIIQ 240

65 Query: 254 EILKQSF SRIPVYDVDDKDKIIGLIHTKRLLESGRQGFQINMRKMLQEPLFVPETIFVD 313

-1342-

IL + FSRIPVYD DKDKIIG+IHTK LL +GF++GFD IN+R++LQEPLFVPETI V+
 Sbjct: 241 TILNERFSRIPVYDDDKDKIIGIHTKNLLNAGFKEGFDHINLRRLQEPLFVPETIVVN 300

Query: 314 DLLRQLRNTQNQMAILLDEYGGVAGLVTTLEDLLEEIVGEIDDETDKAEQFVHEIGDNTYI 373
 DLL L+NTQNQMAILLDEYGGVAGLVTTLEDLLEEIVGEIDDETDK V EI DNTYI
 Sbjct: 301 DLLTALKNTQNQMAILLDEYGGVAGLVTTLEDLLEEIVGEIDDETDKTAISVREIADNTYI 360

Query: 374 VVGTMTLNEFNDFDTELESDDVDVTIAGFYLTGIGTIPSQEQKEAYEIDNKKHLVLIND 433
 V+GTMTLN+FN+YF+T+LESD+VDTIAGFYLTG+GTIPSQE+KE +E+++ KHL LIND
 Sbjct: 361 VLGTMTLNDNFNEYFETDLESDNVDTIAGFYLTGVTGIPSQEEKEHFEVESNGKHLLELIND 420

Query: 434 KVKDGRITKLLKILSNIEQIIIE 456
 KVKDGR+TKLK+++S +E+ +E
 Sbjct: 421 KVKDGRVTKLKILVSEVEEKEDE 443

An alignment of the GAS and GBS proteins is shown below.

Identities = 364/444 (81%), Positives = 417/444 (92%)

Query: 1 MQDPGSQSLLLQFVILLILTLFNAFFSASEMALVSLNRSKVEQKAEEDKRYRRLLDVLE 60
 M+DP SQSL++QF++L++LTL NAFFSASEMALVSLNRS+VEQKA +GDK+Y RLL VLE
 Sbjct: 14 MEDPVSQSLVIQFLLLVVLTLNNAFFSASEMALVSLNRSRVEQKAADGDKKYARLLRVLE 73

Query: 61 NPNNFLSTIQVGITFISLLQGASLSASLGHVISGWLGN SATARTAGSIIALIFLTYVSIV 120
 PN+FLSTIQVGITFISLL GASLSASLG VISGWLGN SATARTAG+II+L+FLTYVSIV
 Sbjct: 74 EPNHFLSTIQVGITFISLLSGASLSASLGKVISGWLGN SATARTAGTIISLVFLTYVSIV 133

Query: 121 LGELYPKRIAMNLKDRLAIVSAPIIIFLGKIVSPFVWLLSASTNLLSRITPMTFDDADEK 180
 LGELYPKRIAMNLKD+LAIVSAPIII LG++VSPFVWLLSASTNLLSR+TPMTFDDADE+
 Sbjct: 134 LGELYPKRIAMNLKDKLAIVSAPIIIGLGRVSPFVWLLSASTNLLSRITPMTFDDADEQ 193

Query: 181 MTRDEIEYMLTNSEETLEAEEIEMLGIFSLDEMAREVMVPRTDAFMIDINND AQSNIE 240
 MTRDEIEYML+ SE TL+AEEIEMLG+FSLDEMAREVMVPRTDAFMIDIN+D NI+
 Sbjct: 194 MTRDEIEYMLSKSEATLDAEEIEMLGVSFSLDEMAREVMVPRTDAFMIDINDDPLENIQ 253

Query: 241 GILSQNFSRVPVFDDDKDRVGVLTHTKRLLEAGFKTGFDITDLRKILQEPLFVPETIFVD 300
 IL Q+FSR+PV+D DKD+++G++HTKRLLE+GF+ GFD I++RK+LQEPLFVPETIFVD
 Sbjct: 254 EILKQFSRIPVYDVDDDKDKIIGLIHTKRLLESGRQGFQINMRKMLQEPLFVPETIFVD 313

Query: 301 DLLKALRNTQNQMAILLDEYGGVAGLVTTLEDLLEEIVGEIDDETD AEQFVREIDENIYI 360
 DLL+ LRNTQNQMAILLDEYGGVAGLVTTLEDLLEEIVGEIDDETD AEQFV EI +N YI
 Sbjct: 314 DLLRQLRNTQNQMAILLDEYGGVAGLVTTLEDLLEEIVGEIDDETDKAEQFVHEIGDNTYI 373

Query: 361 VLGTMTLNEFNDFEETLESDDVDVTIAGYYLTGVSIPNQEKVAYEVD SKDKHITLIND 420
 V+GTMTLNEFNDFY+TELESDDVDVTIAG+YLTG+G+IP+QE+K AYE+D+KDKH+ LIND
 Sbjct: 374 VVGTMTLNEFNDFDTELESDDVDVTIAGFYLTGIGTIPSQEQKEAYEIDNKKHLVLIND 433

Query: 421 KVKDGRITKLLVLLSDIEQNIEKD 444
 KVKDGRITKLLK++LS+IEQ IE+D
 Sbjct: 434 KVKDGRITKLLKILSNIEQIIIEED 457

SEQ ID 3718 (GBS70d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 120 (lane 8-10; MW 65kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 120 (lane 11 & 12; MW 44kDa) and in Figure 179 (lane 5; MW 35kDa).

GBS70d-His was purified as shown in Figure 231, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1343-

Example 1195

A DNA sequence (GBSx1271) was identified in *S.agalactiae* <SEQ ID 3721> which encodes the amino acid sequence <SEQ ID 3722>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1212(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB84230 GB:AL162754 hypothetical protein NMA0960 [Neisseria meningitidis Z2491]

Identities = 80/184 (43%), Positives = 119/184 (64%), Gaps = 3/184 (1%)

Query: 1 MIKRPIHLSHDFLAVIDKEAITLDATMGNGNDTVFLAKSSK---KVYAFDIQEEAIAKT 57
++K + +H L + + + LD T GNG+DT+FLA+++ KV+AFDIQ +A+ T

Sbjct: 2 LLKNILPFAHCLLRQALPEGGNALDGTAGNGHDTLFLAQTAGIRGKVVAFDIQPQALNNT 61

Query: 58 KAKLTEQGISNAELILDGHENLEQYVHTPLRAAIFNLGYLPSADKTIVITKPHTTIKAIKN 117
+ +L E G SN LILDGHENL+QY+ PL AAIFN G+LP DK++ T+ T+I A+

Sbjct: 62 RCRLQEAGYSNVRLLILDGHENLKQYIPKPLDAAIFNFGWLPGGDKSLTTRTETSIAALSA 121

Query: 118 VLDILEVGGRLSLMVYVYGHGGKSEKDAVIAFVEQLPQNNFATMLYQPLNQVNTPPFLIM 177
L +L+ G L ++Y GH+ GK E +A+ + + LPQ FA + Y N+ N+PP+L+

Sbjct: 122 ALSLLKENGMLIAVLYPGHENGKQEAIEQWAKNLPQEQAFAVLRYSFTNRKNSPPYLLA 181

Query: 178 VEKL 181

EKL

Sbjct: 182 FEKL 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3723> which encodes the amino acid sequence <SEQ ID 3724>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.44 Transmembrane 127 - 143 (123 - 143)

----- Final Results -----

bacterial membrane --- Certainty=0.1574(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9101> which encodes the amino acid sequence <SEQ ID 9102>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.44 Transmembrane 118 - 134 (114 - 134)

----- Final Results -----

bacterial membrane --- Certainty= 0.157(Affirmative) < succ>

bacterial outside --- Certainty= 0.000(Not Clear) < succ>

bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 124/184 (67%), Positives = 156/184 (84%)

Query: 1 MIKRPIHLSHDFLAVIDKEAITLDATMGNGNDTVFLAKSSKVVYAFDIQEEAIAKTKAK 60
M+KRPIHLSHDFLAEV+DK ++ +DATMGNGNDT FLA+ +KKVYAFD+QE+AI KT +

-1344-

Sbjct: 10 MLKRPIHLSHDFLAEVVDKSSVVVDATMGNGNDTAFLAQLAKKVYAFDVQEQAIRKTSER 69

Query: 61 LTEQGISNAELILDGHENLEQVHTPLRAAIFNLGYLPSADKTVITKPHTTIKAIKNVLD 120
L + G+SNAILIL GHE ++QYV P+RAAIFNLGYLPSADK++IT P+TT++A+ +L

5 Sbjct: 70 LAQLGLSNAELILAGHEAVDQYVTEPVRAAIFNLGYLPSADKSIITLPTTLQALSLLT 129

Query: 121 ILEVGGRLSLMVYYGHDGGKSEKDAVIAFVEQLPQNNFATMLYQPLNQVNTPPFLIMVEK 180
+L VGGR+++MVYYGHDGG EKDA++ FV+QL Q + MLYQPLNQVNTPPFLIM+EK

10 Sbjct: 130 LLMVGGRIAIVYYGHDGGSLEKDALLDVFKQLDQKVSAMLYQPLNQVNTPPFLIMLEK 189

Query: 181 LQSY 184
L +

Sbjct: 190 LADF 193

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1196

A DNA sequence (GBSx1272) was identified in *S.agalactiae* <SEQ ID 3725> which encodes the amino acid sequence <SEQ ID 3726>. Analysis of this protein sequence reveals the following:

20 Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1948(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAC00380 GB:AF008220 YtqA [Bacillus subtilis]
Identities = 161/302 (53%), Positives = 220/302 (72%), Gaps = 4/302 (1%)

Query: 2 KKRYRAINDYYRELFGEKIFKLPIDAGFDCPNRDGTVARGGCTFCTVSGSGDAIVAPEAP 61
+KRY +N + RE FG K+FK+ +D GFDCPNRDGTVA GGCTFC+ +GSGD

35 Sbjct: 13 EKRYHTLNYHLREHFGHKVFKVALDGGFDCPNRDGTVAHGGCTFCSAAGSGDFAGNRTDD 72

Query: 62 IREQFYKEIDFMHRKWPEVNKYLVIYFQNFNTNTHAKLEIIKERYEQAINEPGVIGINIGTR 121
+ QF+ + MH KW + KY+ YFQ FTNTHA +E+++E++E + V+GI+I TR

40 Sbjct: 73 LITQFHDIKNRMHEKWKD-GKYIAYFQAFNTNTHAPVEVLREKFESVLALDDVVGISIATR 131

Query: 122 PDCLPDETIYYLAELSERMHVTLLELGLQTTYEATSALINRAHSYDLYKKTVKRIRELAPK 181
PDCLPD+ + YLAEL+ER ++ +ELGLQT +E T+ LINRAH ++ Y + V ++R+

45 Sbjct: 132 PDCLPDDVDYLAELNERTYLVWELGLQTVHERTALLINRAHDFNCYVEGVNKLKHHG-- 189

Query: 182 VEIVSHLINGLPGETHDMMVENVRRCTVDNDIQGIKLHLLHLLMTNTRMQRDYHEGRLRL 241
+ + SH+INGLP E DMM+E + V D D+QGIK+HLLHL+ T M + Y +G+L L

50 Sbjct: 190 IRVCSHIINGLPLED RMMMETAK-AVADLDVQGIKIHLHLLKGTMPVKQYEKGKLEFL 248

Query: 242 SQEDYISIIICDQLEIIPKHIVIHRTGDAPRHMLIGPMWSLNKWEVLNAIDKEMEKRSY 301
SQ+DY+ ++CDQLEIIP +++HRITGD P ++IGPMWS+NKWEVL AI+KE+E R SY

55 Sbjct: 249 SQDDYVQLVCDQLEIIPPEMIVHRITGDGPIELMIGPMWSVNKWEVLGAINKELENRGSY 308

Query: 302 QG 303
QG

Sbjct: 309 QG 310

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3727> which encodes the amino acid sequence <SEQ ID 3728>. Analysis of this protein sequence reveals the following:

Possible site: 57
>>> Seems to have no N-terminal signal sequence

60

-1345-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2023(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 260/307 (84%), Positives = 290/307 (93%), Gaps = 1/307 (0%)

Query: 1 MKKRYRAINDDYYRELFGEKIFKLPIDAGFDCPNRDGTVARGGCTFCTVSGSGDAIVAPEA 60
 MKKRY+ +N++YR+LFG K+FK+PIDAGFDCPNRDGTVA GGCTFCTVSGSGDAIVAP+A
 Sbjct: 7 MKKRYQTLNEHYRQLFGAKMFKVPIDAGFDCPNRDGTVAHGGCTFCTVSGSGDAIVAPDA 66

Query: 61 PIREQFYKEIDFMHRKWPEVNKYL VYFQNFNTNTHAKLEIKERYEQAINEPGVIGINIGT 120
 PI+EQFYKEIDFMHRKW+VN+YLVYFQNFNTNTH +++I++RYEQAINEPGV+GINIGT
 Sbjct: 67 PIKEQFYKEIDFMHRKWPDVNRYL VYFQNFNTNTHDVTVDVIRDRYEQAINEPGVGINIGT 126

Query: 121 RPDCLPDETIYYLAELSERMHVTLELGLQTTYEATSALINRAHSYDLYKKTVKRIRELAP 180
 RPDCLPD+TI YLAELSERMHVT+ELGLQTTYE TS LINRAHSYDLYK+TV+R+R P
 Sbjct: 127 RPDCLPDDTIAYLAELSERMHVTVELGLQTTYEETSRLINRAHSYDLYKETVRRLRHY-P 185

Query: 181 KVEIVSHLINGLPGETHDMMVENVRRCTVDNDIQGIKLHLLHMTNTRMQRDYHEGRLRL 240
 + IVSHLINGLP ETHDMM+ENVRRCTVDNDIQGIKLHLLHMTNTRMQRDYHEGRL+L
 Sbjct: 186 NINIVSHLINGLPKETHDMMLENVRRCTVDNDIQGIKLHLLHMTNTRMQRDYHEGRLKL 245

Query: 241 LSQEDYISIIICDQLEIIPKHIVHRITGDAPRHMLIGPMWSLNKWEVLNAIDKEMEKRQS 300
 LSQ+DY+SIICDQLEIIPKHIVHRITGDAPR MLIGPMWSLNKWEVLNAIDKEME+R S
 Sbjct: 246 LSQKDYVSIICDQLEIIPKHIVHRITGDAPRMLIGPMWSLNKWEVLNAIDKEMERRGS 305

Query: 301 YQGCKAE 307
 +QGCK +
 Sbjct: 306 FQGCKVD 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1197

A DNA sequence (GBSx1273) was identified in *S. agalactiae* <SEQ ID 3729> which encodes the amino acid sequence <SEQ ID 3730>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.82	Transmembrane	10 - 26 (6 - 30)
INTEGRAL	Likelihood = -4.73	Transmembrane	93 - 109 (87 - 112)
INTEGRAL	Likelihood = -4.57	Transmembrane	163 - 179 (161 - 181)
INTEGRAL	Likelihood = -2.97	Transmembrane	189 - 205 (185 - 205)
INTEGRAL	Likelihood = -1.97	Transmembrane	58 - 74 (58 - 74)
INTEGRAL	Likelihood = -0.75	Transmembrane	130 - 146 (130 - 146)

----- Final Results -----

bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA79986 GB:Z21972 ORF2 [Bacillus megaterium]

Identities = 62/159 (38%), Positives = 92/159 (56%), Gaps = 3/159 (1%)

Query: 34 ISFDQTIQESVRGQLPNLSTRFFKLITVIGNTVSGIAIAIMSVTFY--LKKWYPQARFI 91
 + FD+ + V+G L T K T IG+T S I ++++ + F Y LK F
 Sbjct: 34 LKFDDEDVISLVQGWESPLLTDIMKFYTYIGSTASLIILSLVILFFLYRILKHRLELVLEFT 93

Query: 92 AVNAIISGICILSLKLIFQVRPTLTHLVFAGGYSFPGHSMGTFMIFGSIILLQYYMP 151
 AV + S + L +KL FQR RP L L+ GYSFPGSH+M F ++G + LL ++

-1346-

Sbjct: 94 AV-MVGSPLNLMLVVKLFFQRRPDLHRLIDIGGYSFPSGHAMNAFSLYGILTFLLWRHIT 152

Query: 152 KSIWKLLCQGTGLGLLIFLIGLSRIYLGVHFPDVLGFI 190

++L L+I IG+SRIYLGVH+P+D++AG++

Sbjct: 153 ARWARILLILFSLMILSIGISRIYLGVHYPDIAGYL 191

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1851> which encodes the amino acid sequence <SEQ ID 1852>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.30 Transmembrane 154 - 170 (150 - 181)

INTEGRAL Likelihood = -10.88 Transmembrane 65 - 81 (58 - 93)

INTEGRAL Likelihood = -8.97 Transmembrane 10 - 26 (5 - 31)

INTEGRAL Likelihood = -3.77 Transmembrane 86 - 102 (86 - 105)

INTEGRAL Likelihood = -2.71 Transmembrane 185 - 201 (183 - 202)

INTEGRAL Likelihood = -1.54 Transmembrane 130 - 146 (130 - 148)

----- Final Results -----

bacterial membrane --- Certainty=0.5522(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/197 (44%), Positives = 134/197 (67%), Gaps = 1/197 (0%)

Query: 1 MLSRQNSKLIQAFIAIILFFSLGLVIKYWPDVVISFDQTIQESVRGQLPNLSTRFFKLIT 60

M ++Q LI +F A+++F +G +K++P+ + D TIQ +RG LP + T+FF+ +T

Sbjct: 2 MTNKQTHFLIASF-ALLIFVIIGYTVKFFPERLALLDNTIQAEIRGNLPVLTQFFRGVT 60

Query: 61 VIGNTVSQIAIAIMSVTFCYLKKWYPQARFIAVNAIISGICILSLKLIFQVRPTLTHLV 120

V GN ++Q+ + I+SV + KW +A FI N I+ I +LKL +QR RP + HL V

Sbjct: 61 VFGNVMTQVLLVIVSVLVLFFMKWKIEALFILSNGAIAAFLITTLKLFYQRPRAIEHLV 120

Query: 121 FAGGYSFPSGHSMTFMIFGSIILLQYMPKSIWKLLCQGTGLGLLIFLIGLSRIYLGVH 180

+AGGYSFPSGH+MG+ +IFGS++I+ + + + + +LI LIGLSRIYLGVH

Sbjct: 121 YAGGYSFPSGHAMGSMILIFGSLIICYQLHSLKLQFVTSMIFIILLIGLSRIYLGVH 180

Query: 181 FPTDVLGFIAYGILN 197

+P+D+LAGF+L +GIL+

Sbjct: 181 YPSDILAGFVLGFGILH 197

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1198

A DNA sequence (GBSx1274) was identified in *S.agalactiae* <SEQ ID 3731> which encodes the amino acid sequence <SEQ ID 3732>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -8.44 Transmembrane 35 - 51 (33 - 59)

INTEGRAL Likelihood = -6.53 Transmembrane 193 - 209 (179 - 211)

INTEGRAL Likelihood = -4.46 Transmembrane 64 - 80 (60 - 82)

INTEGRAL Likelihood = -4.09 Transmembrane 108 - 124 (103 - 128)

INTEGRAL Likelihood = -2.71 Transmembrane 150 - 166 (148 - 166)

INTEGRAL Likelihood = -0.06 Transmembrane 174 - 190 (174 - 190)

----- Final Results -----

bacterial membrane --- Certainty=0.4376(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1347-

A related GBS nucleic acid sequence <SEQ ID 9977> which encodes amino acid sequence <SEQ ID 9978> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5      >GP:AAC83944 GB:L47648 putative [Bacillus subtilis]
      Identities = 53/186 (28%), Positives = 109/186 (58%)

      Query: 33  RKMVTIAILSALSFLMMVSFPLIPGAFLKVDLSILPMLVAFILFDLKSSYGVLLLRSL 92
                  +K+V +++LS+++FVLM+++FP      ++LK+DFS +P ++A +++  + V ++++
10     Sbjct: 4   KKLIVVVSMLSSIAFVLMMLNFPFPGLEPDYLLKIDFSDVPAIIAILIYGPLAGIAVEAIKNV 63

      Query: 93  LKVILANRGPETFIGLPMNMVALALFLASFAIFWKNRESAKDFIKASLFGTVSLTVSMVA 152
                  L+ I+      +G N +A LF+  A +K  SAK  + L GT ++T+ M
      Sbjct: 64  LQYIIQGSMAGVPVGQVANFIAGTLFILPTAFLFKKLNSAKGLAVSLLLGTAAMTILMSI 123

15     Query: 153 LNVVFAIPLAIFANFDIRTFIVGNVLLTMVIPFNIVEGILISIVFYLTIVVACLPIER 212
                  LNVV +P Y F +  +  +  ++ ++PFN+++GI+I++VF L ++ P +E+
      Sbjct: 124 LNVVLILPAYTWFLHSPALSDSALKTAVVAGILPFNMIGIVITVVFSLIFIKLKPWIEQ 183

      Query: 213 YKKTINV 218
                  +  ++
20     Sbjct: 184 QRSAHI 189
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3733> which encodes the amino acid sequence <SEQ ID 3734>. Analysis of this protein sequence reveals the following:

```

25     Possible site: 26
      >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -6.48    Transmembrane    82 - 98 ( 74 - 100)
      INTEGRAL    Likelihood = -3.93    Transmembrane    161 - 177 ( 152 - 178)
30     INTEGRAL    Likelihood = -3.61    Transmembrane    108 - 124 ( 107 - 126)
      INTEGRAL    Likelihood = -3.61    Transmembrane    33 - 49 ( 31 - 50)

      ----- Final Results -----
                  bacterial membrane --- Certainty=0.3590(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the databases:

```

40     >GP:AAC83944 GB:L47648 putative [Bacillus subtilis]
      Identities = 46/182 (25%), Positives = 97/182 (53%)

      Query: 3   KTHKMIMIGILSAISFLMLVVSFAIIPGAFLKIEFSIIPVLFGLMIMDLKSAYLILLR 62
                  K K++++ +LS+I+F+LML++F      +LKI+FS +P +  ++I  +  +  ++
      Sbjct: 2   KVKKLIVVVSMLSSIAFVLMMLNFPFPGLEPDYLLKIDFSDVPAIIAILIYGPLAGIAVEAIK 61

45     Query: 63  SLLKLFLNNRGVNDFIGLPMNIIAIALFVTAFAFVWNRQKTLQYVFASLLGTGLLTFGM 122
                  ++L+ +      +G N IA LF+  A ++ +  +  + LLGT +T M
      Sbjct: 62  NVLQYIIQGSMAGVPVGQVANFIAGTLFILPTAFLFKKLNSAKGLAVSLLLGTAAMTILM 121

      Query: 123  VVLNYTFAIPLAIFANIDIRAYIGVTKYMMTMVIPFNLVEGLIFAITFYFVYIASKPIL 182
                  +LNY +P Y F +  +  +  ++ ++PFN+++G++ + F ++I KP +
50     Sbjct: 122 SILNYVLILPAYTWFLHSPALSDSALKTAVVAGILPFNMIGIVITVVFSLIFIKLKPWI 181

      Query: 183  ER 184
                  E+
55     Sbjct: 182  EQ 183
  
```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 110/185 (59%), Positives = 144/185 (77%)

60     Query: 29  MTNTRKMVTIAILSALSFLMMVSFPLIPGAFLKVDLSILPMLVAFILFDLKSSYGVLL 88
                  M+ T KM+ I ILSA+SF+LM+VSF +IPGA FLK++FSI+P+L  ++ DLKS+Y +LL
  
```

-1348-

Sbjct: 1 MSKTHKMIMIGILSAISFLLMLVVSFAIIPGAAFLKIEFSIIPVLFGLMIMDLKSAYLILL 60

Query: 89 LRSLLK+ L NRG FIGLPMN++A+ALF+ +FA+ W +++ ++ ASL GT LT 148

5 Sbjct: 61 LRSLLKLFNNRGVNDFIGLPMNIIAIALFVTAFALVWNRQKTLQYVFASLLGTGLLTF 120

Query: 149 SMVALNYVFAIPLYAIFANFDIRTFIGVGNVLLTMVIPFNIVEGILISIVFYLTIVACLP 208

MV LNY FAIPLYAIFAN DIR +IGV Y++TMVIPFN+VEG++ +I FY Y+A P

10 Sbjct: 121 GMVVLNYTFAIPLYAIFANIDIRAYIGVTKYMMTMVIPFNLVEGLIFAITFYFVYIASKP 180

Query: 209 ILERY 213

ILERY

Sbjct: 181 ILERY 185

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1199

A DNA sequence (GBSx1275) was identified in *S.agalactiae* <SEQ ID 3735> which encodes the amino acid sequence <SEQ ID 3736>. Analysis of this protein sequence reveals the following:

20 Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -11.04 Transmembrane 278 - 294 (270 - 298)

----- Final Results -----

25 bacterial membrane --- Certainty=0.5416(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 30 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3736 (GBS150) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 7; MW 29.7kDa) and in Figure 175 (lane 4 & 5; MW 30kDa).

Purified GBS150-His is shown in Figure 110A, Figure 199 (lane 5) and Figure 227 (lanes 6-7).

- 35 The purified GBS150-His fusion product was used to immunise mice (lane 1+2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 110B), FACS (Figure 110C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1200

A DNA sequence (GBSx1276) was identified in *S.agalactiae* <SEQ ID 3737> which encodes the amino acid sequence <SEQ ID 3738>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

45 Possible site: 40

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -15.34 Transmembrane 264 - 280 (257 - 285)

INTEGRAL Likelihood = -7.64 Transmembrane 23 - 39 (12 - 41)

----- Final Results -----

-1349-

bacterial membrane --- Certainty=0.7135(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 95/271 (35%), Positives = 139/271 (51%), Gaps = 16/271 (5%)

10 Query: 29 VGLLITSYPFISNWYYNIKANNQVTNFDNQTKLNTKEINRRFELAKAYNRTLDP SRLSD 88
 +GLL +YP ++W + ++ Q + + E A AYN L + +
 Sbjct: 1 MGLL--TYPTAASWVSQYNQSKVTADYSAQVDGARP-DAKTQVEQAHAYNDALSAGAVLE 57

15 Query: 89 PYTE-----KEKKGIAEYAHMLEIAE--MIGYIDIPSIKQKLPYAGTTSSVLEKGAGH 140
 K +YA++L+ ++ + IPSI LP+Y GT L KG GH
 Sbjct: 58 ANNHVPTGAGSSKDSLLQYANILKANNEGLMARLKIPSISLDLPVYHGTADDTLKGLGH 117

20 Query: 141 LEGTSLPIGGKSSHTVITAHRLPKAKLFTDLKLLKKGKIFYIHNIKEVLAYKVDQISVV 200
 LEGTSLP+GG+ + +VIT HRGL +A +FT+LDK+K G + EVL Y+V VV
 Sbjct: 118 LEGTSLPVGGEGTRSVITGHRGLAEATMFTNLDKVKTGDSLIVEVFGEVLTYRVTSTKVV 177

Query: 201 KPDNFSKLLVVKGDYATLLTCTPYSINSHRLVRGHRKIYVPPVKEKNYLMKELQTHYK 260
 +P+ L V +GKD TL+TCTP IN+HR+L+ G RI Y P K+ K +
 Sbjct: 178 EPEETEALRVEEGKDLLTLVTCTPLGINTHRILLTGERI-YPTPAKDLAAAGKRPDVPHP 236

25 Query: 261 LYFLLSILVILILVALLL----YLKRKFKEK 287
 ++ + + LI+V L L Y + KER
 Sbjct: 237 PWWAVGLAAGLIVVGLYLWRSYGAAARAKER 267

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3739> which encodes the amino acid sequence <SEQ ID 3740>. Analysis of this protein sequence reveals the following:

Possible site: 49

35 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-14.01 Transmembrane 225 - 241 (220 - 248)

----- Final Results -----
 bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 45 Identities = 94/250 (37%), Positives = 133/250 (52%), Gaps = 17/250 (6%)

Query: 1 VECYRDRQLLSTYHKQVTQKKPSEMEEVWQKAKAYNARLGIQVPDAF-----SFRD 52
 V Y ++ + Y QV +P +V ++A AYN L V +A S +D
 Sbjct: 13 VSQYNQSKVTADYSAQVDGARPDAKTQV-EQAHAYNDALSAGAVLEANNHVPTGAGSSKD 71

50 Query: 53 GIHDKNYESLLQIENNDIMGYVEVPSIKVTLPIYHYTTDEVLTGAGHLFGSALPVGGDG 112
 Y ++L+ N +M +++PSI + LP+YH T D+ L KG GHL G++LPVGG+G
 Sbjct: 72 S--SLQYANILKANNEGLMARLKIPSISLDLPVYHGTADDTLKGLGHLEGTSLPVGGEG 129

55 Query: 113 THTVISAHRLPSAEMFTNLNLVKKGDTFYFRVLNKLAYKVDQILTVEPDQVTSLSGVM 172
 T +VI+ HRGL A MFTNL+ VK GD+ V +VL Y+V VEP++ +L
 Sbjct: 130 TRSVITGHRGLAEATMFTNLDKVKTGDSLIVEVFGEVLTYRVTSTKVVPEETEALRVEE 189

60 Query: 173 GKDYATLVTCTPYGVNTRKLLVRGHRIAHYKQYQAKKAMKLVDKSRMWAEEVVCFAFGV 232
 GKD TLVTCTP G+NT R+L+ G RI Y K + K A G+
 Sbjct: 190 GKDLLTLVTCTPLGINTHRILLTGERI-----YPTPAKDLAAAGKRPDVPHPWWAVGL 243

Query: 233 VIAIILVFMV 242
 +I+V +Y

-1350-

Sbjct: 244 AAGLIVVGLY 253

An alignment of the GAS and GBS proteins is shown below.

Identities = 93/192 (48%), Positives = 130/192 (67%), Gaps = 2/192 (1%)

```

5  Query: 52  VTNFDNQTKLNTKEINRRFELAKAYNRTLDP SRLSDPYTEKEKKGIAEYAHMLEIA--E 109
      ++ + Q + E+ ++ AKAYN L + D ++ ++ Y +L+I +
Sbjct: 10  LSTYHKQVTQKKPSEMEEVWQKAKAYNARLGIQVPDAFSFRDGIHDKNYESLLQIENND 69

10 Query: 110 MIGYIDIPSIKQLPIYAGTTSSVLEKGAGHLEGTSLPIGGKSSHTVITAHRG LPKAKLF 169
      ++GY+++PSIK LPIY TT VL KGAGHL G++LP+GG +HTVI+AHRG L P A++F
Sbjct: 70  IMGYVEVPSIKVTLPIYHYTTDEVLTGAGHLFGSALPVGGDGTHTVISAHRGLPSAEMF 129

15 Query: 170 TDLCLKKKGIIFYIHNIKEVLAYKVDQISVVKPDNFSKLLVVGKDYATLLTCTPYSINS 229
      T+L+ +KKG FY + +VLAYKVDQI V+PD + L V GKDYATL+TCTPY +N+
Sbjct: 130 TNLNLVKKGDTFYFRVLNKLVLAYKVDQILTVEPDQVTSLSGVMGKDYATLVTCTPYGVNT 189

Query: 230 HRLLVRGHRIKY 241
      RLLVVRGHRI Y
20 Sbjct: 190 KRLLVRGHRIAY 201

```

SEQ ID 3738 (GBS210) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 3; MW 61kDa).

GBS210d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 2-4; MW 54kDa) and in Figure 187 (lane 9; MW 54kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 2-4; MW 28.7kDa) and in Figure 182 (lane 13; MW 29kDa). Purified GBS210d-GST is shown in lane 4 of Figure 237.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1201

A DNA sequence (GBSx1277) was identified in *S.agalactiae* <SEQ ID 3741> which encodes the amino acid sequence <SEQ ID 3742>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

```

35 Possible site: 42
    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL    Likelihood = -10.61    Transmembrane    20 - 36 ( 15 - 40)
        INTEGRAL    Likelihood = -7.27     Transmembrane    259 - 275 ( 258 - 277)

40 ----- Final Results -----
        bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

45 >GP:AAC13546 GB:AF019629 putative fimbria-associated protein
    [Actinomyces naeslundii]
    Identities = 76/219 (34%), Positives = 120/219 (54%), Gaps = 12/219 (5%)

50 Query: 28  LSILLYPVVSRFFYYTIESNNQTDQFERAAKLSQKEINRRMALAQAYNDSL N-----N 80
      + +L YP + + + T D+ A ++ + ++ A AYND+L+ N
Sbjct: 1  MGLLTYP TAA SWVSQYNQSKVTADYS-AQVDGARPD AKTQVEQAHAYNDALSAGAVLEA N 59

Query: 81  VHLEDPYEKKRIQKGVAEYARMLLEVSEK--IGTISVPKIGQKLPIFAGSSQEVLSKGAGH 138

```


-1352-

```

: | : | : | | | : | | | | | | : | | | : | | : : :
EVFGEVLTYRVTSTKVVEPEETEALRVEEGKDLLTLVTCTPLGINTHRIILLTGERI-YPTPAKD-LAAAGKRPDVPHPFW
      170      180      190      200      210      220      230

5  1098      1179      1209      1239      1269      1299
Y-----ISLVIIAWLLWL--IKRQRQKNR-LASVRKGIES*WEENFRKTLRNRSF*IDG*M*A*YYCS*LVF**PHILF
: | : : | | | | | | | | : : | | : : | | :
WAVGLAAGLIVVGLYLWRSGYAAARAKERALARARAAQEPPQQTWAEQMRIWMDDDAGVEPQRWFTDLFPVPQPSEMEN
      250      260      270      280      290      300      310

```

SEQ ID 8750 (GBS212) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 4; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 2; MW 61kDa).

Purified Thio-GBS212-His is shown in Figure 244, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1202

A DNA sequence (GBSx1278) was identified in *S.agalactiae* <SEQ ID 3743> which encodes the amino acid sequence <SEQ ID 3744>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL      Likelihood =-10.40      Transmembrane      680 - 696 ( 674 - 699)

----- Final Results -----
bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA57459 GB:X81869 orf2 [Lactobacillus leichmannii]
Identities = 84/325 (25%), Positives = 122/325 (36%), Gaps = 94/325 (28%)

Query: 397 VNVVYTLKDKD-----KTVASVSLTKTSKGTI---DLGNGIKFEVSGNF 437
      VNV + +KDKD      TV+ LTK++ T+ D G + F+ +
Sbjct: 236 VNVFWNKDKDKDTFNVVDKPDGTGIDIDASTVSDGLTKSTDYTVNKKDNGYQVVFKTT--- 292

Query: 438 SGKFTGLENKSYMISERVSGYGSAINLENGKVITINTKSDNPTPLNPTEPKVETHGKKF 497
      S L KS I+      K T+TN D + T +G
Sbjct: 293 SAAVQALAGKSLTITY-----KATLTNNATPDKA--IGNATATLSIGNGTNI 336

Query: 498 VKTNEQGDRLL--AGAQFVVKNSAGKYLALKADQSEGQKTLAAKKIALDEAIAAYNKLSAT 555
      T G R+ GAQFV K+S      + KTLA + L + + N +S
Sbjct: 337 TSTPANGPRIYTGAQFVKKDS-----QSNKTLAGAEFQLVKVDSNGNIVSYA 384

Query: 556 DQKGEKGITAKELIKTKQADYDAAFIEARTAYEWITDKARAITYSNDQGQFEVTGLADG 615
      Q +      +Y W A TYTS+ G + GL+
Sbjct: 385 TQASDG-----SYTWNDSATEATYTS DANGLVALKGLSYS 420

Query: 616 -----TYNLEETLAPAGFAKLAGNIKFFVNVQGSYITGGNIDYVANSNQKDATTRVENKK 668
      +Y L E AP G+AKL +KF + QGS+ G+ + + N K+
Sbjct: 421 DKLDGSGSYALLEIQAPDGYAKLDSPVKFSITQGSF---GDSNKITIDNTKEG----- 470

Query: 669 VTIPQTGGIGTILFTIIGLSIMLGA 693
      +P TGG G +F IG+ IM+ A
Sbjct: 471 -LLPSTGGKGIYIFLAIGIVIMIVA 494

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-1353-

SEQ ID 3744 (GBS59) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 8; MW 120kDa), in Figure 11 (lane 9; MW 100kDa) and in Figure 13 (lane 6; MW 74kDa).

GBS59-His was purified as shown in Figure 193, lane 2.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1203

A DNA sequence (GBSx1279) was identified in *S.agalactiae* <SEQ ID 3745> which encodes the amino acid sequence <SEQ ID 3746>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 25
    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL    Likelihood = -3.13    Transmembrane  870 - 886 ( 864 - 887)

    ----- Final Results -----
15         bacterial membrane --- Certainty=0.2253(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP:AAD33086 GB:AF071083 fibronectin-binding protein I [Streptococcus pyogenes]
    Identities = 58/176 (32%), Positives = 83/176 (46%), Gaps = 19/176 (10%)

    Query: 6   KFSKILTLSLFCLSQIPLNTNVLGEST--VPENGA--KGKLVVKKTDQNKPLSKATFV 60
              K S +L+L+ F L + + + G S      NGA +G +KK D NKPL AT
25 Sbjct: 8   KLSFLLSLTGFI LGLLLVF IGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATSS 67

    Query: 61  LKTTAHPESKIEKVTAELTGEATFDNLIPGDYTLSEETAPEGYKKTNTQTWQVKVESNGKT 120
              L +      + ++ T+   G      NL PG YTL EETAP+GY KT++TW V V NG T
30 Sbjct: 68  LTSKDGGKTSVQFTTSNDKGIVDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYT 127

    Query: 121 TIQNSGDKNSTIGQNQEELDKQYPPTGIYEDTKESYKLEHVKGSVPN--GKSEAKA 174
              + +      I +      +D   S +Le+ K SV + GK+E +
35 Sbjct: 128 KLVENFPYNGEIIISKAGS-----KDVSSSLQLENPKMSVSVSKYGKTEVSS 171
    Identities = 31/92 (33%), Positives = 49/92 (52%), Gaps = 14/92 (15%)

    Query: 725 PTITIKNEKKLGEIEFIKVDKDNKLLKLGATFELQEFNEDYKLYLPIKNNNSKVVTGEN 784
              P+IT+ N K++ ++ F K+ DN + L A FEL+ N      N+ K+ N
40 Sbjct: 501 PSITVANLKRVAQLRFKIMSTDN--VLPLEAAAFELRSSN-----GNSQKLEASSN 548

    Query: 785 --GKISYKDLKDGKYQLIEAVSPEDYQKITNK 814
              G++ +KDL G Y L E +P+ YQ++T K
40 Sbjct: 549 TQGEVHFKDLTSGTYDLYETKAPKGYQQVTEK 580

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 SEQ ID 3746 (GBS67) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 10; MW 140kDa), in Figure 11 (lane 10; MW 150kDa) and in Figure 12 (lane 6; MW 95.3kDa).

GBS67-His was purified as shown in Figure 192, lane 10.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1354-

Example 1204

A DNA sequence (GBSx1280) was identified in *S.agalactiae* <SEQ ID 3747> which encodes the amino acid sequence <SEQ ID 3748>. This protein is predicted to be Nra. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2020(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9979> which encodes amino acid sequence <SEQ ID 9980> was also identified.

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3749> which encodes the amino acid sequence <SEQ ID 3750>. Analysis of this protein sequence reveals the following:

```

    Possible site: 58
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.75    Transmembrane    393 - 409 ( 392 - 409)
20
    ----- Final Results -----
      bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 122/325 (37%), Positives = 186/325 (56%), Gaps = 5/325 (1%)

30 Query: 7  LIENYLEKDILNQIKLLTLCY--DYYPISITLDKSCHQLGLSELLIRKYCHDLTTLFNSQL 64
    LIE YLE I ++ +L+ L + Y P + + + GL+ L + YC +L F L
    Sbjct: 1  LIEKYLESSIESKQCQLIVLFFKTSYLP---ITEVAEKTGLTFLQLNHYCEELNAFFPGSL 57

    Query: 65  SLNIEKSTIVYQSNQVGTREQAFKYIYHQSHVLQLLKFLITNDSGRLPLTYFSEKFGLSCA 124
    S+ I+K I Q +E +Y S+VLQLL FLI N S PLT F+ LS +
35 Sbjct: 58  SMTIQKRMI SCQFTHPFKETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNS 117

    Query: 125  TAYRIRKHISPLLEKLGFIQIVKNTITGDEYRIRYLI AFLNAQFGIEVYPMSKMDKLLIKR 184
    +AYR+R+ + PLL ++ KN I G+EYRIRYLI A L ++FGI+VY +++ DK I
40 Sbjct: 118  SAYRMREALIPLLRNFELKLSKNKIVGEEYRIRYLIALLYSKFGIKVYDLTQQDKNTIHS 177

    Query: 185  LLEHSTTFTASHYFPNTFIFFDTLLSLSWKRINYNVVPYSSLFTELQNIIFYDTLQYC 244
    L ST S + +F F+D LL+LSWKR ++V +P + +F +L+ +F+YD+L+
    Sbjct: 178  FLSSHSTHLKTSFPLSESFYDILLALSWKRHQFSVTIPQTRIFQQLKKLFVYDSLKKS 237

45 Query: 245  VKNVIIDSFKINLKKDDIDYIFLAYLTSHNSFSNPNWTEKRIDNVIAIFENYPKFQKLLQ 304
    ++I ++N D+DY++L Y+T++NSF++ WT + I +FE F+ LL
    Sbjct: 238  SHDIETFCQLNFSAGDLDYLYLIYITANNSFASLQWTPHIRQYCYQLFEENDTFRLLLN 297

    Query: 305  PLKDALPLSGSYHDELVKVAIFFSE 329
    P+ LP LVK +FFS+
50 Sbjct: 298  PIITLLPNLKEQKASLVKALMFFSK 322

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1355-

Example 1205

A DNA sequence (GBSx1281) was identified in *S.agalactiae* <SEQ ID 3751> which encodes the amino acid sequence <SEQ ID 3752>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 21
   >>> Seems to have no N-terminal signal sequence (or aa 1-22)

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.1168(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAB99071 GB:U67549 galactosyltransferase isolog [Methanococcus
      jannaschii]
      Identities = 108/395 (27%), Positives = 196/395 (49%), Gaps = 28/395 (7%)

Query: 4   KVKTVAVFSGYYLPFLGGIERYTDKMTADLVK-RGYRVVIVTTHGDLPIIDEDKGR--- 59
          K+K + +F GYY+P +GG+E + D+ T L + .Y + I N +P E + R
20   Sbjct: 3   KIKLI-IFPGYIPHIGGLETHVDEFTKHLSEDENYDIYIFAPN---IPKYKEFEIRHNN 58

Query: 60   -KIYRLPTKNIVKQRYPIINK-NREYNTLMKYVSDENIDFVICNTRFQLTTLGLELSFAKN 117
          K+YR P I+ YP+ N N ++ + + + D V+ TRF TL G FAK
25   Sbjct: 59   VKVYRYPAFEIIPN-YPVPNIFNIKFWRMFFNLYKIDFDIVMTRTRFFSNTLLGFIFAKL 117

Query: 118   HHLPS--IVLDHGSSHSFVNNRFLDFFGAIYEHLLTARVKHYRPDFYAVSKRSVEWLKHF 175
          I ++HGS+ + + F + Y+ + + A+SK ++
30   Sbjct: 118   RFKKKKLIHVEHGSFAFKLESEFNKLSYFYDKTIGKLIFKKADYVVAISKAVKNFILEN 177

Query: 176   NIEAKGV--IYNSVS---ESLGSDFAGTAYLEKSADDIFITYAGRIIKEKGIELLLEAF 229
          + K + IY + ES+G D EK + I + + GR+ K KG+E +++A+
35   Sbjct: 178   FVNDKDIPIIYRGLIEKIESIGED---KKIKEKFKNKIKLCFVGRLYKWKGVENIIKAY 234

Query: 230   S--MSQYSENVYLQIAGDGPPELAHLKE---KYQSKQINFLGKLNFEQTMSLMAQTDFIVY 284
          E + L + G G +L LK+ Y + I F GK++FE+ ++++ +DI+++
40   Sbjct: 235   VDLPKDLKEKIILIVVGYGEDLERLKKLAGNYLNNGIYFTGKVDFEKAIAIVKASDIYIH 294

Query: 285   PSMYPEGLPTSILEAGLLSSAIIATDRGGTVEVIDSPELGIIMEENT-QLSHESLDLLVK 343
          S GL +S+L+A AI+A+ G EV+ GI++++N+ + + + L++
45   Sbjct: 295   SSYKGGGLSSSLQAMCCGKAIVASPYEGADEVVIDGYNGILKDNSPEEIKRGIKILIE 354

Query: 344   DKALREKLQONIAKRIKEHFTWEKTVEKLDYIIQK 378
          + LR+ +N IKE+F W+K+V++ I ++
50   Sbjct: 355   NNNLRKIYGENAKNFIKENFNWKKSVKEYKKIFER 389

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3752 (GBS258) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 2; MW 43kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 7; MW 67.9kDa).

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1206

A DNA sequence (GBSx1282) was identified in *S.agalactiae* <SEQ ID 3753> which encodes the amino acid sequence <SEQ ID 3754>. Analysis of this protein sequence reveals the following:

```

55   Possible site: 31

```

-1356-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1182(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB52237 GB:Z98171 EpsQ protein [Streptococcus thermophilus]
 Identities = 112/278 (40%), Positives = 163/278 (58%), Gaps = 2/278 (0%)

Query: 1 MKYLAGIVTFNPNIERLDQNIIRAIYPQVSHIYIVDNGSKNKEEISQLVADYNEEGHLTVD 60
 M AGIV FNP+I+RL +NI A+ Q +H+Y+VDNGS N +E+ L+ YN+ +++
 Sbjct: 1 MDISAGIVLFPNDIKRLKENIDAVIIQCTHLYLVLDNGSGNVDEVKGLLNQYNQS-KISIL 59

15 Query: 61 YLTENKGIAYALNCIGQFAVAQEFDWFLTLQDQSVVLGDLIDNYENYLHLPKVGMLSCLY 120
 + EN+GIA ALN + A + FDW LTLQDQSVV +++ +E Y++ VG+L +
 Sbjct: 60 WNRENQGIKALNQLTSAAQKEGFDWILTLDQDQSVVPSNIVGEFEKYINNSSVGILCPII 119

20 Query: 121 QDMNRENLMQEFDYKEIEECITSAAALMKTSVFEETSGFAEEMFIDFVDSEMNRYLSEMG 180
 D N++ + D EI+ECITS +L+ + E GF E MFID VD ++ YRL + G
 Sbjct: 120 CDRNKDEBIKINEDCTEIDECITSGSLLNIKAWSEIGGFDERMFIDGVDFDICRYLRQRG 179

25 Query: 181 YKTYQVNFIFGLLHEIGHSSRVKKFGHVHVLNHSPPFRKYYMIRNAIYIICKYGKKKRYKY 240
 YK Y ++ + LLHE+GH + V NHS FRKYY+ RN IY KK
 Sbjct: 180 YKIYCIHSVVLHLELGHIEYHRFLFWKVLVKNHSAFRKYYIARNIYITAKKRRSTLLVVK 239

Query: 241 LVFMRNEFVRVLV-AEEQKSKKIVAMIKGLKDGLLMKV 277
 + + + +++ EE K KI + +G+ DG KV
 30 Sbjct: 240 GLLQEIKLIGIVIFYEEDKLNKIRCICRGIYDGFKGKV 277

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 1207

A DNA sequence (GBSx1283) was identified in *S.agalactiae* <SEQ ID 3755> which encodes the amino acid sequence <SEQ ID 3756>. This protein is predicted to be EpsU protein (rfbX). Analysis of this protein sequence reveals the following:

40 Possible site: 54
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.44	Transmembrane	357 - 373 (352 - 387)
INTEGRAL	Likelihood = -7.59	Transmembrane	88 - 104 (79 - 107)
INTEGRAL	Likelihood = -7.32	Transmembrane	440 - 456 (433 - 465)
INTEGRAL	Likelihood = -6.48	Transmembrane	246 - 262 (245 - 263)
45 INTEGRAL	Likelihood = -4.78	Transmembrane	294 - 310 (290 - 312)
INTEGRAL	Likelihood = -3.88	Transmembrane	164 - 180 (162 - 183)
INTEGRAL	Likelihood = -3.56	Transmembrane	144 - 160 (136 - 161)
INTEGRAL	Likelihood = -2.87	Transmembrane	317 - 333 (316 - 334)
INTEGRAL	Likelihood = -2.71	Transmembrane	374 - 390 (374 - 393)
50 INTEGRAL	Likelihood = -0.96	Transmembrane	44 - 60 (44 - 62)
INTEGRAL	Likelihood = -0.80	Transmembrane	15 - 31 (15 - 32)

----- Final Results -----

55 bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB52225 GB:Z98171 EpsU protein [Streptococcus thermophilus]

-1357-

Identities = 189/462 (40%), Positives = 313/462 (66%)

Query: 1 MKLLKNMFYNTSYQLLTLLPLVTPVYVSRVLSPPQGINAYTSSIVMYFTLFGALGISL 60
 M+++KN YN YQ+ +++PL+T+PY+SR+L P GIGIN+YT+SIV YF LFG++G+ L
 5 Sbjct: 1 MQIVKNYLYNAIQVFIIIVPLLTIPYLSRILGPSGIGINSYTNISIVQYFVLFSGISGLGL 60

Query: 61 YGNREIAFVQSNKYKRSKIFWELVVLKLASVSIATLLFFGFVLLTNEWQLFYLIQGINLL 120
 YGNR+IAFV+ N+ K SK+F+E+ +L+L ++ +A LF F+++ ++ +YL Q I ++
 10 Sbjct: 61 YGNRQIAFVRDNQVKMSKVFEIFILRLFTICLAYFLFVAFPLIINGQYYAYYLSQSIIV 120

Query: 121 ATATDISWYFIGVEDFKIIVIRNTIVKLITVVLTFVLVVKTPDDLALYMFLIAFASLLGNL 180
 A A DISW F+G+E+FK+IV+RN IVKL+ + FL VK+ +DL +Y+ + ++L+GNL
 15 Sbjct: 121 AAADFISWAFMGIEFVKVIVLRNFIVKLLALFSIFLKVSYNDLNIYILITVLSLTLGNL 180

Query: 181 TVWHHLKHEIIPKPSRLDILHLRPTMLFLPQITMQIYLSLNKSMGLGAMDSVVSAGYF 240
 T + L ++K+ + L + HL+ +L++F+PQI +QIY LNK+MLG++DSV S+G+F
 20 Sbjct: 181 TFFPSLHRYLVKVNRYELRPIKHLKQSLVMFIPQIALQIYVVLNKTMLGSLDSVTSSGFF 240

Query: 241 DQSDKIIRILFTIVSAIGGVFLPRLSSLFSSGKEKQAKALLKLVDLSNAISMLMIAGVV 300
 DQSDKI++++ IV+A G V LPR+++ F+ + + K + +AIS+ M+ G++
 25 Sbjct: 241 DQSDKIVKLVLAIVTATGTVMLEPRVANAFHREYSKIKEYMYAGFSFVSAISIPMMFGLI 300

Query: 301 GVSSTFAVFFPGKGYEAVGPLMAVESLMIICISYGNALGTQYLLASRRTKAYTMSAVIGL 360
 ++ F FF + V P++ +ES+ II I++ NA+G QYLL + + K+YT+S +IG
 30 Sbjct: 301 AITPKFVPLPFTSQFSDVIPVLMIESIAIIFIAWSNAIGNQYLLPTNQNKSYSYTVSVIIGA 360

Query: 361 VANVVLNILLIPILGAMGAIISTVITEFIVSLYQAIISLRDVFTFKELTRGMLRYLIAATL 420
 + N++LNI LI LGA+GA I+TVI+E V++YQ + L + +YLIA +
 35 Sbjct: 361 IVNLMLNIPLIYILGAVGASIAATVISEMSVTYVYQLFIIHKQLNLHTLFSDLISKYLIAGLV 420

Query: 421 SGAVLYYINTQMSVSLVNYVIQSLVAVTIYVGIVFITKAPVI 462
 +++ I+ S + +++ V + IY+ ++ KA +I
 40 Sbjct: 421 MFLIVFKISLLTPTSWIFILEITVGIIYIVLLIFLKAEII 462

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1208

40 A DNA sequence (GBSx1284) was identified in *S.agalactiae* <SEQ ID 3757> which encodes the amino acid sequence <SEQ ID 3758>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.1742(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1209

55 A DNA sequence (GBSx1285) was identified in *S.agalactiae* <SEQ ID 3759> which encodes the amino acid sequence <SEQ ID 3760>. Analysis of this protein sequence reveals the following:

-1358-

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

10 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1210

15 A DNA sequence (GBSx1286) was identified in *S.agalactiae* <SEQ ID 3761> which encodes the amino acid sequence <SEQ ID 3762>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

20 INTEGRAL Likelihood = -10.56 Transmembrane 214 - 230 (210 - 236)
 INTEGRAL Likelihood = -10.03 Transmembrane 364 - 380 (361 - 386)
 INTEGRAL Likelihood = -7.96 Transmembrane 272 - 288 (271 - 291)
 INTEGRAL Likelihood = -6.95 Transmembrane 23 - 39 (20 - 41)
 INTEGRAL Likelihood = -5.57 Transmembrane 191 - 207 (189 - 209)
 INTEGRAL Likelihood = -5.15 Transmembrane 434 - 450 (425 - 451)
 INTEGRAL Likelihood = -4.25 Transmembrane 143 - 159 (138 - 162)
 25 INTEGRAL Likelihood = -3.13 Transmembrane 167 - 183 (166 - 186)
 INTEGRAL Likelihood = -1.44 Transmembrane 400 - 416 (400 - 416)
 INTEGRAL Likelihood = -1.33 Transmembrane 333 - 349 (333 - 349)
 INTEGRAL Likelihood = -0.80 Transmembrane 232 - 248 (232 - 251)

30 ----- Final Results -----

 bacterial membrane --- Certainty=0.5225 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1211

40 A DNA sequence (GBSx1287) was identified in *S.agalactiae* <SEQ ID 3763> which encodes the amino acid sequence <SEQ ID 3764>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.1792 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50

A related GBS nucleic acid sequence <SEQ ID 9981> which encodes amino acid sequence <SEQ ID 9982> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAF18951 GB:AF155805 Cps9H [Streptococcus suis]
    Identities = 53/116 (45%), Positives = 75/116 (63%), Gaps = 4/116 (3%)

    Query: 6  VLMATYNGQGFIHDQLDSIRNQTLRPDYVLMRDDGSTDDTVKVVEDYIKEHRLDGWSITS 65
              VLMATYNG FI QLDSIRNQ++ D V++ DD STDDT+K+++DYIK++ LD W ++
    10  Sbjct: 4  VLMATYNGSPFIKQLDSIRNQSVSADKVIWDDCSTDDTIKIKDYIKKYSLDSWVVSQ 63

    Query: 66  NDKNLGWRNLNFRQLLIDVLAYEVDYVFFSDQDDTWYHHKNMQVDIMEERQDINLL 121
              N N G F L + VFFSDQDD W HK + + I +R+++++
    10  Sbjct: 64 NKSNGQHYQTFINL---TKLVQEGIVFFSDQDDIWDCHKIETMLPIF-DRENVSMV 115

```

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1212

20 A DNA sequence (GBSx1288) was identified in *S.agalactiae* <SEQ ID 3765> which encodes the amino acid sequence <SEQ ID 3766>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

```

    Possible site: 13
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    25  bacterial cytoplasm --- Certainty=0.1278(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 A related GBS nucleic acid sequence <SEQ ID 9983> which encodes amino acid sequence <SEQ ID 9984> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAF18951 GB:AF155805 Cps9H [Streptococcus suis]
    Identities = 57/146 (39%), Positives = 81/146 (55%), Gaps = 8/146 (5%)

    35  Query: 10  VLMATYNGEIFISEQLDSIRQQTLKPDYVLLRDDCSTDETVNVNNYIAKHELEGWKIVK 69
              VLMATYNG FI +QLDSIR Q++ D V++ DDCSTD+T+ ++ +YI K+ L+ W + +
    Sbjct: 4  VLMATYNGSPFIKQLDSIRNQSVSADKVIWDDCSTDDTIKIKDYIKKYSLDSWVVSQ 63

    40  Query: 70  NDKNLGWRNLNFRQLLIDVLAYEVDYVFFSDQDDIWLKNERQFAIMSDKPQIEVLSADV 129
              N N G F L + VFFSDQDDIW K E I D+ + + V
    Sbjct: 64 NKSNGQHYQTFINL---TKLVQEGIVFFSDQDDIWDCHKIETMLPIF-DRENVSM---V 115

    Query: 130 DIKTMSTEASVPFLTFSSSDRISQY 155
              K+ + + + +SDRI+ Y
    45  Sbjct: 116 FCKSRLIDENGNISSPDTSDRINTY 141

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1213

50 A DNA sequence (GBSx1289) was identified in *S.agalactiae* <SEQ ID 3767> which encodes the amino acid sequence <SEQ ID 3768>. This protein is predicted to be dTDP-glucose 4-6-dehydratase (galE). Analysis of this protein sequence reveals the following:

-1360-

Possible site: 44

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.02 Transmembrane 250 - 266 (250 - 266)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1808(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 A related GBS nucleic acid sequence <SEQ ID 9985> which encodes amino acid sequence <SEQ ID 9986> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC14890 GB:AJ295156 d-TDP-glucose dehydratase [Phragmites
australis]

15 Identities = 108/327 (33%), Positives = 170/327 (51%), Gaps = 22/327 (6%)

Query: 29 ANKGVLLISGSNSMLASYMVFLLAYLNETRNYQTQIIATARNIEKARDKFSDDLVGKDYFTL 88

AN +L++G + S++V L N + ++I ++D +G F L

20 Sbjct: 33 ANLRILVTGGAGFIGSHLVDKLM-----ENEKHEVIVADNFFTGSKDNLKKWIGHPRFEL 87

Query: 89 IPYDVEERLEYDGKVDYIIHAASNASPTAILSNPVSIIKANTIGTLNLLDFAKEKTIENF 148

I +DV + L + VD I H A ASP NPV IK N IGTLN+L AK +

Sbjct: 88 IRHDVTQPLLVE--VDQIYHLACPASPIFYKHNPVKTIKTNVIGTLNMLGLAK-RVGARI 144

25 Query: 149 LFLSTREVGTSIKEVIDEEAYGGFDILATRACYPESKRMAETLLQSYDQYKVPFTIAR 208

L ST EVYG ++ E +G + + R+CY E KR+AETL+ Y+ Q+ + IAR

Sbjct: 145 LLTSTSEVYGDPLEHPQTEAYWGNVNPVGVRSCYDEGKRVAETLMFDYHRQHGIEIRIAR 204

30 Query: 209 IAHSFGPGMELGNDGRIMNDLLSNVIDGKDIVLKSSGTAERAFCYLADAVSGLFTILLNG 268

I +++GP M + +DGR++++ ++ + G + ++ GT R+FCY+AD V GL L+NG

Sbjct: 205 IFNTYGPGRMNI-DDGRVVSNFIAQAVRGDPLTVQKPGTQTRSFYVADMVDGLIK-LMNG 262

Query: 269 EVGQAYNVANEDQPIMIKDLAQKLVDLFSKDNISVVFDIPKTMASAGYSKMGRT---LTM 325

N+ N + M+ +LA+K+ +L + ++ TM+ R R +T

35 Sbjct: 263 NNTGPINLGNPGEFTML-ELAEVKELINP-----EVTVTMTENTPDDPRQRKPDITK 314

Query: 326 AKLEALGWKREVSLESGLIKTVQAFEE 352

AK E LGW+ +V L G++ F E

40 Sbjct: 315 AK-EVLGWEPKVVLRLDGLVLMEDDFRE 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1214

45 A DNA sequence (GBSx1290) was identified in *S.agalactiae* <SEQ ID 3769> which encodes the amino acid sequence <SEQ ID 3770>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have an uncleavable N-term signal seq

50 ----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9987> which encodes amino acid sequence <SEQ ID 9988> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11866 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]

-1361-

Identities = 77/231 (33%), Positives = 131/231 (56%), Gaps = 6/231 (2%)

Query: 13 VIFAGGVGRRMNTKGKPKQFLEVHGKPIIVHTIDIFQNTAIDAVVVVCSVDWLDYMNLL 72
 VI A G G+RM G+ K F+E+ G P+I+HT+ +F + D +++V ++ L
 5 Sbjct: 6 VIPAAGQGKRMKA-GRNKLFIELKGDPIIHTLRVFDShRQCDKIILVINEQEREHFQQL 64

Query: 73 VERFNLTKVKAVVAGGETGQMSIFKGLEAAEQLATDDAVVLIHDGVRPLINEEVINANIQ 132
 + + +VAGG+ Q S++KGL+A +Q + +VL+HDG RP I E I+ I
 10 Sbjct: 65 LSDYPFQTSIELVAGGDERQHSVYKGLKAVKQ----EKIVLVHDGARPFIKHEQIDELIA 120

Query: 133 SVKETGSAVTSVRAKETVVLVNDSSKISEVVDRTSRFIAKAPQSFYLSDILSVERDAISK 192
 ++TG+A+ +V K+T+ V D ++SE ++R+ + + PQ+F LS ++ +A K
 Sbjct: 121 EAEQTGAAILAVPVKDTIKRVQDL-QVSETIERSSSLWAVQT'PQAFRLSLLMKAHAAEAERK 179

Query: 193 GITDAIDSSTLMGMYNRELITVEGYPYENIKITTPDDFYMFKALYDARENEQ 243
 G D+S + M + +VEG Y NIK+TTPDD +A+ ++ +
 15 Sbjct: 180 GFLGTDDASLVEQMEGGSVRVVEGSYTNIKLITPDDLTSAAEIMESSESGNK 230

No corresponding DNA sequence was identified in *S.pyogenes*.

20 SEQ ID 3770 (GBS647) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 9 & 10; MW 55.9kDa + lane 8; MW 27kDa) and in Figure 186 (lane 5; MW 56kDa).. It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 12; MW 31kDa), in in Figure 140 (lane 9; MW 31kDa) and in Figure 178 (lane 6; MW 31kDa).

25 Purified GBS647-GST is shown in Figure 243, lane 4; purified GBS647-His is shown in Fig.229, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1215

30 A DNA sequence (GBSx1291) was identified in *S.agalactiae* <SEQ ID 3771> which encodes the amino acid sequence <SEQ ID 3772>. This protein is predicted to be LicD1. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2647(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9989> which encodes amino acid sequence <SEQ ID 9990> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD37094 GB:AF106539 LicD2 [Streptococcus pneumoniae]
 Identities = 85/271 (31%), Positives = 130/271 (47%), Gaps = 15/271 (5%)
 45 Query: 1 MKEMTVSEIREVQLEMLAYIDKVARDNKIEYSLGGGSLGAMRHKGFI PWDDDDIDLMLER 60
 M+ + EI+E+QL +L YID+ + + I Y L G++LGA+RHKG IPWDDDDID+ L R
 Sbjct: 1 MQYLEKKKEIKEIQALLDYIDETCKKHDI PYFLSYGTMLGAIRHKGMI PWDDDDIDISLYR 60

50 Query: 61 SQYERLMKALADANNNSDFKLLHHSVEKNLW---PFAKLYHTKSMYLSKTDRIHPWTGIFI 117
 YERL+K + + N+ +K+L S + + W FA + T ++ T +FI
 Sbjct: 61 EDYERLLKIEEENHPRYKVL--SYDTSSWYFHNFA SILDTS'VIEDHVKYKRHDTSLFI 118

-1362-

Query: 118 DIFPLDRLPESAEERQRFKKVHSAANLMCTTYPNFASGSRKLYANARLILGLP-RFIA 176
 D+FP+DR + + + + + A L G KL RL RF+
 Sbjct: 119 DVFPIDRFTDLSIVDKSY---KYVALRQLAYIKKSRVHGDGSKLKDFLRLCSWYALRFVN 175

5 Query: 177 YHQAKIKRAEIVDQVMETYNQEVFPYMGYTD-SRYRLKEYFPREIFSEYEDVMFENIKTR 235
 KK +DQ+++ Y G + +KE FP + F E FE
 Sbjct: 176 PRYFYKK---IDQLVKNAVNTNPQYEGGVGIGKEGMKEIFPVDTFKELILTEFEGRMLP 231

10 Query: 236 KIKNEHAYLNQLYGGSYMELPPESKRESHSY 266
 K +L Q+Y G YM P + +E +S+
 Sbjct: 232 VPKKYDQFLTQMY-GDYMTPPSKEMQEWYSH 261

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 15 vaccines or diagnostics.

Example 1216

A DNA sequence (GBSx1292) was identified in *S.agalactiae* <SEQ ID 3773> which encodes the amino acid sequence <SEQ ID 3774>. Analysis of this protein sequence reveals the following:

20 Possible site: 18
 >>> May be a lipoprotein
 INTEGRAL Likelihood =-12.05 Transmembrane 554 - 570 (547 - 575)

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 SEQ ID 3774 (GBS182d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 184 (lane 8; MW 62kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1217

35 A DNA sequence (GBSx1293) was identified in *S.agalactiae* <SEQ ID 3775> which encodes the amino acid sequence <SEQ ID 3776>. Analysis of this protein sequence reveals the following:

40 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4653(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1363-

Example 1218

A DNA sequence (GBSx1294) was identified in *S.galactiae* <SEQ ID 3777> which encodes the amino acid sequence <SEQ ID 3778>. This protein is predicted to be DOLICHYL-PHOSPHATE MANNOSE SYNTHASE RELATED PROTEIN. Analysis of this protein sequence reveals the following:

```

5      Possible site: 29
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL    Likelihood = -2.92    Transmembrane 232 - 248 ( 231 - 248)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.2168(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

15 A related GBS nucleic acid sequence <SEQ ID 9991> which encodes amino acid sequence <SEQ ID 9992> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:AAC35924 GB:AF071085 putative glycosyl transferase [Enterococcus
          faecalis]
      Identities = 118/240 (49%), Positives = 152/240 (63%), Gaps = 1/240 (0%)
20      Query: 14  KILLVIPAYNEEGSIKTVQTIIVDFKASRS-LPFELDYIVINDGSTDGTPELLDRLGLNH 72
          K+LL+IPAYNEE +I +T+ +I FK + ELDY+VINDGSTDGT ++L+ +N
      Sbjct: 2    KVLIIIPAYNEENILRTIASIETFKQEVTHFQHELDYVVINDGSTDGTQKILEVNQINA 61

      Query: 73  IDLVQNLGIGGCVQTGYLYANRNHYDVAVQFDGDGQHDIRSIEDVVMPIINDEADDFVIGS 132
          I LV NLGIGG VQTGY YA N YDVA QFDGDG HDI S+ ++ P+ F GS
      Sbjct: 62  IHLVLNLGIGGAVQTGYKYALENEYDVAXQFDGDGXHDIXSLPILLEPLAEGXCXFSXGS 121

      Query: 133 RFVDKKHQNFQSTAMRRGLINLISAAIKLTGCHKVYDTTSGYRAANAALIAYLSCHYFPVQ 192
          RF+ +FQS MRR GI L+S G +Y T G RA N +IA+ + YP
30      Sbjct: 122 RFIPGNXASFQSKMRRXGIRLLSFCXXXAXGXTIYXVTXGXRAGNRKVIAFFAKRYPTN 181

      Query: 193 YPEPESTARILKKGYRLKEVTANMFEREAGTSSISSLSKIFYMTDVLTSIIIAGFIKEDD 252
          YPEPES ++KK + + E NM ER G SSI +L S+ YM +V ++I+IA F+KE D
35      Sbjct: 182 YPEPESIVHLIKKRFVIVERPVNMMERLGGVSSIRALASVKYMLEVGSAILIAPFMKEGD 241

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3779> which encodes the amino acid sequence <SEQ ID 3780>. Analysis of this protein sequence reveals the following:

```

      Possible site: 56
40      >>> Seems to have no N-terminal signal sequence
          INTEGRAL    Likelihood = -0.80    Transmembrane 211 - 227 ( 211 - 227)

      ----- Final Results -----
45      bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

      >GP:AAC35924 GB:AF071085 putative glycosyl transferase [Enterococcus
          faecalis]
      Identities = 104/233 (44%), Positives = 134/233 (56%), Gaps = 9/233 (3%)
50      Query: 1  VKKLIIPAYNESSNIVNTIRTIESDAPD-----FDYIIIDDCSTDNTLAICQKQGFN 53
          +K L+IIPAYNE NI+ TI +IE+ + DY++I+D STD T I + N
      Sbjct: 1  MKVLIIPAYNEENILRTIASIETFKQEVTHFQHELDYVVINDGSTDGTQKILEVNQIN 60

      Query: 54  VISLPINLGIGGAVQTGYRYAQRQGYDVAVQVDGDGQHNP CYLEKMVEVLVQSSVNMVIG 113
          I L +NLGIGGAVQTGY+YA YDVA Q DGDG H+ L ++E L + G
55      Sbjct: 61  AIHLVLNLGIGGAVQTGYKYALENEYDVAXQFDGDGXHDIXSLPILLEPLAEGXCXFSXG 120

```

```

Query: 114 SRFI--TKEGFQSSSFARRIGIKYFTWLIALLTGKKITDTSGLRLIDRSlierFANHYPD 171
          SRFI      FQS  RR GI+  ++      G  I   T G R  +R +I  FA  YP
Sbjct: 121 SRFIPGNXASFQSKMRRXGIRLSFCXXXAGXTIYXVTXGXGRAGNRKVIAFFAKRYPT 180

Query: 172 DYPEPETVDVLDVLSHFVKVKEIPVMMNERQGGVSSISLTSKVYMIKVTILAILV 224
          +YPEPE++V ++'  F + E PV M ER GGVSSI  SV YM++V AIL+
Sbjct: 181 NYPEPEsIVHLIKKRFVIVERPVNMMERLGGVSSIRALASVKYMLEVGSAILI 233

```

Identities = 105/231 (45%), Positives = 142/231 (61%), Gaps = 8/231 (3%)

```

Query: 14 KILLVIPAYNEEGSIAKTVQTVDFKASRSLPFELDYIVINDGSTDGTPELLDRLGLNHI 73
          K L++IPAYNE +I T++TI S + DYI+I+D STD T + + G N I
15 Sbjct: 2 KKLIIIPAYNESSNIVNTIRTI-----ESDAPDFDYIIIDDCSTDNTLAICQKQGFNVI 55

Query: 74 DLVQNLGIGGCVCVTGYLYANRNHYDVAVQFDGQDGHDIRSIEDVMPILNDEADVFVIGSR 133
          L NLGIGG VQTGY YA R YDVAVQ DGDGQH+ +E +V ++ + VIGSR
20 Sbjct: 56 SLPINLGIGGAVQGTGYRYAQRCGYDVAVQVDGQDQHNP CYLEK MVEVLVQSSVMVIGSR 115

Query: 134 FVDKKHQNFQSTAMRRLGINLISAAIKLITGHKVYDTTSGYRAANAALIAYLSCHYPVQY 193
          F+ K + FQS+ RR+GI + I L TG K+ D TSG R + +LI + HYP Y
Sbjct: 116 FITK--EGFQSSFARRIGIKYFTWLIALLTGKKITDATSGRLRIDRSLIERFANHYPDDY 173

Query: 194 PEPESTARILKKGYRLKEVTANMFEREAGTSSISLKSIFYMTDVLTSIII 244
          PEPE+ +L +++KE+ M ER+ G SSIS KS++YM V +I++
25 Sbjct: 174 PEPEVTVDVLVSHFKVEIPVMNERGGVSSISLTKSVYMIKVTLLIIV 224

```

```

Lipop: Possible site: -1    Crend: 9
McG: Discrim Score:      0.29
GvH: Signal Score (-7.5): -4.34
    Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
ALOM program    count: 1 value: -2.92 threshold: 0.0
    INTEGRAL      Likelihood = -2.92    Transmembrane  222- 238( 221 - 238)
    PERIPHERAL    Likelihood = 4.40      4
modified ALOM score: 1.08

```

```

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.2168(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

50 ORF00548(340 - 1056 of 1359)
GP|3608398|gb|AAC35924.1||AF071085(2 - 241 of 241) putative glycosyl transferase
{Enterococcus faecalis}
%Match = 24.7
%Identity = 49.2 %Similarity = 64.2
Matches = 118 Mismatches = 85 Conservative Sub.s = 36

[illegible]

-1365-

```

INDGSTDGTKQILEVNQINAIHLVLNLGIGGAVQTGYKYALENEYDVAXQFDGDGXHDIKSLPILLEPLAEGXCXFSXGS
      60      70      80      90     100     110     120

726      756      786      816      846      876      906      936
5  RFVDKKHQNFQSTAMRRRLGINLISAAIKLTTGHKVYDITTSYRAANAALIAVLSCHYPVQYPEPESTARILKKGYRLKEV
   ||:      :||| ||| || |:|      | :| | | | | :||:: || ||||| :||| : : |
RFIPGNXASFQSKMRRXGIRLLSFCXXXAGXTIYXVTXGXRAGNRKVIAFFAKRYPTNYPEPESTIVHLIKKRFVIVER
      140      150      160      170      180      190      200

10  966      996      1026      1056      1086      1116      1146      1176
   TANMFEREAGTSSISSLSIFYMTDVLTSIIAGFIKEDDK*V*HCKLKCLF*PLSYFI*L*EWLIKTHFLNLVLYLGY*
   || || | ||| :| |: || :| ::|:| | |:| |
PVMMERLGGVSSIRALASVKYMLEVGSAILIAPFMKEGD
      220      230      240

15

```

SEQ ID 8752 (GBS355) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 4; MW 27kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 7; MW 52kDa).

GBS355-GST was purified as shown in Figure 213 (lane 4) and in Figure 216 (lane 6).

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1219

A DNA sequence (GBSx1295) was identified in *S.agalactiae* <SEQ ID 3781> which encodes the amino acid sequence <SEQ ID 3782>. Analysis of this protein sequence reveals the following:

```

25  Possible site: 19
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -1.91    Transmembrane    185 - 201 ( 185 - 201)

30  ----- Final Results -----
        bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

35  >GP:BAA32090 GB:AB010970 rhamnosyltransferase [Streptococcus mutans]
    Identities = 181/315 (57%), Positives = 244/315 (77%), Gaps = 7/315 (2%)

Query: 1  MKVNILMATYNGEKFLAQQIESIQKQTFKEWNLLIRDDGSSDKTCDIIRNFTAKDSRIRF 60
Sbjct: 1  MKVNILMATYNG++F+AQQI+SIQKQTF+ WNLIRDDGSSD T II +F D+RIRF

40  Query: 61  INENEHHNLGVIKSFFTLVNVEVADFYFFSDQDDVWLPEKLSVSLEAAKHKASDVPLLVY 120
      IN ++ N GVIK+F+TL+ YE AD+YFFSDQDDVWLP+KL ++L + + + +PL+VY
Sbjct: 61  INADKRENFVGIKNFYTLTKYEKADYFFSDQDDVWLPQKLELTLASVEKENNQIPLMVY 120

45  Query: 121  TDLKVVNQELNILQDSMIRAQSHHANTTLLPELTENTVTGGTMMINHALAEKW-FTPNDI 179
      TDL VV+++L +L DSMI+ QSHHANT+LL ELTENTVTGGTMM+NH LA++W +D+
Sbjct: 121  TDLTVVDRDLQVLHDSMIKTQSHHANTSLEELTENTVTGGTMMVNHCLAKQWKQCYDDL 180

50  Query: 180  LMHDWFLALLAASLGEEIYYLDLPTQLYRQHNNVLGARTMDKRFK-ILREGPKSIFTRYW 238
      +MHDW+LALLAASLG++IYLD T+LYRQH++NVLGART KR K LR P + +YW
Sbjct: 181  IMHDWYLALLAASLGKLIYLDETTELYRQHESNVLGARTWSKRLKNWLR--PHRLVKKYW 238

55  Query: 239  KLIHDSQKQASLIVDKYGDIMTANDLELIKCFIKIDKQPFMTRLRWLWKYGYSKNQFKHQ 298
      L+ SQ+QAS +++ D+ AN +I+ ++ + Q F+ R++WL +YG++KN+ H
Sbjct: 239  WLVTSSQQQASHLLEL--DLPAANK-AIRAYVTLLDQSFLNRIKWLKQYGFKNRAFHT 295

Query: 299  VVFKWLIATNYYNKR 313

```

-1366-

VFK LI T + +R
 Sbjct: 296 FVFKTLIITKFGYRR 310

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 817> which encodes the amino acid
 5 sequence <SEQ ID 818>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1980(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 178/314 (56%), Positives = 232/314 (73%), Gaps = 6/314 (1%)

Query: 1 MKVNIIMATYNGEKFLAQQIESIQKQTFKEWNLLIRDDGSSDKTCDIIRNFTAKDSRIRF 60
 M +NIL++TYNGE+FLA+QI+SIQ+QT +W LLIRDDGS+D T DIIR F +D RI++
 Sbjct: 1 MNINILLSTYNGERFLAEQIQSIQRQTVNDWTLIRDDGSTDGTQDIIRTFVKEDKRIQW 60

20 Query: 61 INENEHNLGVIKSFFTLVNYEVADFYFFSDQDDVWLPEKLSVS-LEAAKHKASDVPLLV 119
 INE + NLGVIK+F+TL+ ++ AD YFFSDQDD+WL KL V+ LEA KH+ + PLLV
 Sbjct: 61 INEGQTENLGVIKNFYTLKKHQKADVFFSDQDDIWLNDKLEVTLLAEQKHEMT-APLLV 119

25 Query: 120 YTDLKVVNQELNILQDSMIRAQSHHANTTLLPELTENTVTGGTMMINHALAEKWFTPNDI 179
 YTDLKVV Q L + DSMI+ QS HANT+LL ELTENTVTGGTMMI HALAE+W T + +
 Sbjct: 120 YTDLKVVQHLAVCHDSMIKTQSGHANTSLLQELTENTVTGGTMMITHALAEWTTCDGL 179

30 Query: 180 LMHDWFLALLAASLGEIYYLDLPTQLYRQHDNNVLGARTMDKRFKILREGPKSIFTRYWK 239
 LMHDW+LALLA+++G+++YLD+PT+LYRQHD NVLGART KR K P + +YW
 Sbjct: 180 LMHDWYLALLASAIGKLVLVDIPTELYRQHDANVLGARTWSKRMKNWLT-PHHLVKNKYWW 238

Query: 240 LIHDSQKQASLIVDKYGDIMTANDLELIKCFIKIDKQPFMTRLRLWLWKYGYSKNQFKHQV 299
 LI SQKQA L++D + ND EL+ ++ + PF RL L +YG+ KN+ H
 35 Sbjct: 239 LITSSQKQAQLLLDL---PLKPNDELHVTAYVSLDMPFTKRLATLKRYGFRKNRIFHTF 295

Query: 300 VFKWLIATNYYNKR 313
 +F+ L+ T + +R
 40 Sbjct: 296 IFRSLVVTLFGYRR 309

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1220

A DNA sequence (GBSx1296) was identified in *S.agalactiae* <SEQ ID 3783> which encodes the amino
 45 acid sequence <SEQ ID 3784>. This protein is predicted to be rgpAc. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.1881(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9993> which encodes amino acid sequence <SEQ ID 9994> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1367-

>GP:BAA32089 GB:AB010970 rgpAc [Streptococcus mutans]
Identities = 234/362 (64%), Positives = 284/362 (77%)

5 Query: 33 VSELINHQKSFEDIKYHVACLSDKEHHTHFNADADCFTINPPQLGPARVIAYDIMAINYA 92
+ EL+ +++S + YHVACLSD + H HF + DCFTI P+LGPARIAYD+MAI YA
Sbjct: 1 MEELVYKQSQQLTYHVACLSETDQKHFTYLGVDCTTIKAPKLGPARVIAYDMMAIRYA 60

10 Query: 93 LDLVKTHDLKEPIFYILGNTIGAFIWHFANKIHKVGGLLYVNPDLGLEWKRKWSRPTQRY 152
L L+K +K PIFYILGNTIGAF+ FA KI ++GG Y+NPDGLEW+RSKWSRP Q Y
Sbjct: 61 LKLIKQDKIKHPIFYILGNTIGAFMGPFARKIKRIGGRFYINPDGLEWRRSKWSRPFVQAY 120

15 Query: 153 LKYAEKCMTKNADLIISDNIGIENYIQSTYSNVKTRFIAYGTEINSRKLSSDDPRVKQLF 212
LKYAEKCMTK ADL+ISDN GIE YI+ Y KT FIAYGT+++ L +D +VK +
Sbjct: 121 LKYAEKCMTKKADLVISDNTGIEGYIKQMPWAKTTFIAYGTDLSPLGLKNDKSKVDFY 180

20 Query: 213 KKWNISKGYLIVGRFVPENNYETAIREFMASDTKRDLVIICNHQNNPYFEKLSLKTNL 272
KKW IK KGYLIVGRFVPENNYETAIREFM S ++RDLVIICN++ N YFE L KT
Sbjct: 181 KKWAIKDKGYLIVGRFVPENNYETAIREFMTSSSERDLVIICNYEGNAYFEDLRQKTEF 240

25 Query: 273 QQDKRVKFGVGTLYEKDLLDYVRQQAFAIYIHGHEVGGTNPGLLEALANTDLNLVLDVDFNK 332
+DKR+KFGVT+Y++ LL Y+R+QAFAYIHGHEVGGTNPGLLEALA+TDLNLVL +FN
Sbjct: 241 DKDKRIKFGVTYDRPLTYIREQAFAYIHGHEVGGTNPGLLEALAHTDLNLVLITEFNY 300

30 Query: 333 SVAGLSSFYWAKKEGDLAKLINDSDQQDLSTYGDRAKAIQENYTWKKIVEEYEDLFLN 392
+VA ++ YW + G LA+LIN D+Q++ + YG RAK II YTW+KIVEEYEDLFL+
Sbjct: 301 TVALDAARYWTQDNGSLAQLINQFDKQENFAEYGGRAKEIIVNYTWKIVEEYEDLFLH 360

Query: 393 ES 394
ES
Sbjct: 361 ES 362

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3785> which encodes the amino acid sequence <SEQ ID 3786>. Analysis of this protein sequence reveals the following:

35 Possible site: 23
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.38 Transmembrane 95 - 111 (95 - 111)
----- Final Results -----
40 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 250/383 (65%), Positives = 307/383 (79%)

45 Query: 11 MQDVFIIGSRGLPARYGGFETFVSELINHQKSFEDIKYHVACLSDKEHHTHFNADADCFT 70
MQDVFIIGSRGLPA+YGGFETFV ELI+HQ S +I+YHVACLSD +H HF++ ADCF
Sbjct: 1 MQDVFIIGSRGLPAKYGGFETFVEELISHQSSKNIRYHVACLSDTKHKVHFYKGDADCFY 60

50 Query: 71 INPPQLGPARVIAYDIMAINYALDLVKTHDLKEPIFYILGNTIGAFIWHFANKIHKVGGL 130
+NPP+LGPARIAYD+MAI YAL H ++ PIFY+LGNT+GAFI F +IH GG
Sbjct: 61 LNPPKLGPARIAYDMMAITVALSYSDQHQIQNPIFYVLGNTVGAFIAPFVKQIHNRRGR 120

55 Query: 131 LYVNPDLGLEWKRKWSRPTQRYLKYAEKCMTKNADLIISDNIGIENYIQSTYSNVKTRFI 190
++NPDGLEWKRKWSRP Q YLK++EK MT+ ADL+ISDNIGI+ Y++ Y KT FI
Sbjct: 121 FFINPDGLEWKRKWSRPFVQAYLKFSEKQMTQADLVISDNIGIDRYLKQVYPWSKTCFI 180

60 Query: 191 AYGTEINSRKLSSDDPRVKQLFKWNISKGYLIVGRFVPENNYETAIREFMASDTKR 250
AYGT+ +L++ D +V+ F+ ++I+ K YYLI+GRFVPENNYETAI+EFMAS TKRD
Sbjct: 181 AYGTQTQPSRLATADSKVRAYFQTFDIREKDYLLILGRFVPENNYETAIKEFMASSTKR 240

65 Query: 251 LVIICNHQNNPYFEKLSLKTNLQQDKRVKFGVGTLYEKDLLDYVRQQAFAIYIHGHEVGGTN 310
LVIICNH+ N YF++L +T +D R+KFGVTLY+K+LL Y+R+QA+AYIHGHEVGGTN
Sbjct: 241 LVIICNHEGNAYFKQLLAETECDDPRIKFGVTLYDKELLAYIREQAYAYIHGHEVGGTN 300

-1368-

Query: 311 PGLLEALANTDLNLVLDVDFNKSVAAGLSSFYWAKKEGDLAKLINDSDQQQLSTYGDRAK 370
 PGLLEALA+T+LNLVL VDFN+SVA ++ YW K++G LA+LIN D D G AK
 Sbjct: 301 PGLLEALAHTNLNLVLGVDFNQSVAKSAALYWTQKQQLAELINQVDAGFDSHDLGKEAK 360

5 Query: 371 AIIQENYTWKKIVEEYEDLFLNE 393
 AIIQE+YTW+KIV EYE LFLNE
 Sbjct: 361 AIIQEHYTWKIVGEYEALFLNE 383

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1221

A DNA sequence (GBSx1297) was identified in *S.agalactiae* <SEQ ID 3787> which encodes the amino acid sequence <SEQ ID 3788>. This protein is predicted to be dTDP-L-rhamnose synthase. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1059(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD10184 GB:AF026471 Cps20 [Streptococcus pneumoniae]
 Identities = 258/283 (91%), Positives = 274/283 (96%)

Query: 1 MILITGANGQLGSELRHLLDERTQEYVAVDVAEMDITNAEMVDKVFEEVKPSLVYHCAAY 60
 MILITGANGQLG+ELR+LLDER +EYVAVDVAEMDIT+AEMV+KVFEEVKP+LVYHCAAY
 Sbjct: 1 MILITGANGQLGTELRYLLDERNEEYVAVDVAEMDITDAEMVEKVFEEVKPTLVYHCAAY 60

Query: 61 TAVDAAEDEGKELDFAINVTGTENVAKAAKHDA TLVYISTDYVFDGEKPVGQEWVDDL 120
 TAVDAAEDEGKELDFAINVTGT+NVAKA+ KH ATLVIYISTDYVFDG+KPVGQEWVDD
 Sbjct: 61 TAVDAAEDEGKELDFAINVTGTKNVAKASEKHGATLVYISTDYVFDGKKPVGQEWVDDR 120

Query: 121 PDPKTEYGRTRKRMGEELVEKYTSKFYTIRTAWVFGNYGKNFVFTMQNLAKTHKTLTVVND 180
 PDP+TEYGRTRKRMGEELVEK+ S FY IRTAWVFGNYGKNFVFTMQNLAKTHKTLTVVND
 Sbjct: 121 PDPQTEYGRTRKRMGEELVEKHVSNFYIIRTAWVFGNYGKNFVFTMQNLAKTHKTLTVVND 180

Query: 181 QHGRPTWTRTLAEFMTYLAENQKDFGYHLSNDAKEDTTWYDFAVEILKDTDVEVKPVDS 240
 Q+GRPTWTRTLAEFMTYLAEN+K+FGYHLSNDA EDTTWYDFAVEILKDTDVEVKPVDS
 Sbjct: 181 QYGRPTWTRTLAEFMTYLAENRKEFGYHLSNDATEDTTWYDFAVEILKDTDVEVKPVDS 240

Query: 241 SQFPAKAKRPLNSTMSLEKAKATGFVIPTWQDALKEFYKQEVK 283
 SQFPAKAKRPLNSTMSL KAKATGFVIPTWQDAL+EFYKQEV+
 Sbjct: 241 SQFPAKAKRPLNSTMSLAKAKATGFVIPTWQDALQEFYKQEVK 283

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3789> which encodes the amino acid sequence <SEQ ID 3790>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0618(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 227/284 (79%), Positives = 248/284 (86%)

-1369-

Query: 1 MILITGANGQLGSELRLHLLDERTQEVVAVDVAEMDITNAEMVDKVFEEVKPSLVYHCAAY 60
 MILITG+NGQLG+ELR+LLDER +YVAVDVAEMDITN + V+ VF +VKP+LVYHCAAY
 Sbjct: 21 MILITGSNGQLGTELRYLLDERGVYVAVDVAEMDITNEDKVEAVFAQVKPTLVYHCAAY 80

Query: 61 TAVDAAEDEGKELDFAINVTGTENVAKAAKHDAITLVYISTDYVFDGEKPVGQEWVDDL 120
 TAVDAAEDEGK L+ AINVTG+EN+AKA K+ ATLVYISTDYVFDG KPVGQEW D
 Sbjct: 81 TAVDAAEDEGKALNEAINVTGSENIACGKYGATLVYISTDYVFDGNKPVGQEWVETDH 140

Query: 121 PDPKTEYGRTRKMGEEELVEKYTSKFYTIRTAWVFGNYGKNFVFTMQNLAKTHKTLTVVND 180
 PDPKTEYGRTRK+GE VE+Y FY IRTAWVFGNYGKNFVFTM+ LA+ H LTVVND
 Sbjct: 141 PDPKTEYGRTRKRLGELAVERYAEHFYIIRTAWVFGNYGKNFVFTMEQLAENHSRLTVVND 200

Query: 181 QHGRPTWTRTLAEFMTYLAENQKDFGYHLSNDAKEDTTWYDFAVEILKDTDVEVKPVDS 240
 QHGRPTWTRTLAEFM YL ENQK FGYHLSNDAKEDTTWYDFA EILKD VEV PVDS
 Sbjct: 201 QHGRPTWTRTLAEFMCYLTENQKAFGYHLSNDAKEDTTWYDFAKEILKDKAVEVVPVDS 260

Query: 241 SQFPAKAKRPLNSTMSLEKAKATGFVIPTWQDALKEFYKQEVKK 284
 S FPAKAKRPLNSTM+L+KAKATGFVIPTWQ+ALK FY+Q +KK
 Sbjct: 261 SAFFPAKAKRPLNSTMNLDKAKATGFVIPTWQEALKAFYQQGLKK 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1222

25 A DNA sequence (GBSx1298) was identified in *S.agalactiae* <SEQ ID 3791> which encodes the amino acid sequence <SEQ ID 3792>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2554(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA21508 GB:AB000631 unnamed protein product [Streptococcus mutans]
 Identities = 92/108 (85%), Positives = 100/108 (92%)

Query: 5 KQYSEEEVGKIKDRILEALEMVIDPELGIDIVNLGLIYEIRFEDNGRTEIDMTLTTMGCP 64
 K Y+ EE+ KIKDRILEALEMVIDPELGIDIVNLGLIY+IRFED+GRTEIDMTLTTMGCP
 Sbjct: 4 KNYTPEEIAKIKDRILEALEMVIDPELGIDIVNLGLIYDIRFEDSGRTEIDMTLTTMGCP 63

Query: 65 LADLLTDQIHDVMKTVPEVTETEVKLVWYPAWSVDKMSRYARIALGIR 112
 LADLLTDQIHD +K VPEV + +VKLVW PAW+VDKMSRYARIALGIR
 Sbjct: 64 LADLLTDQIHDALKDVPEVLDIDVWSPAWTVDKMSRYARIALGIR 111

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3793> which encodes the amino acid sequence <SEQ ID 3794>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2818(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 90/112 (80%), Positives = 102/112 (90%)

-1370-

Query: 1 MSEVKQYSEEEVGKIKDRILEALEMVIDPELGIDIVNLGLIYEIRFEDNGRTEIDMTLTT 60
 MS+ +Y+++V IK+RILEALE VIDPELGID+VNLGLIYEIRF DNG TEIDMTLTT
 Sbjct: 1 MSDTPKYTDQVIAIKNRILEALETVIDPELGIDVVNLGLIYEIRFNDNGYTEIDMTLTT 60

5 Query: 61 MGCPPLADLLTDQIHDVMKTVPEVTETEVKLVWYPAWSVDKMSRYARIALGIR 112
 MGCPPLADLLTD IHD ++ VPEVT+TEVKLVWYPAW+VDKMSRYARIALGIR
 Sbjct: 61 MGCPPLADLLTDYIHDALQDVPEVTETEVKLVWYPAWTVDKMSRYARIALGIR 112

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1223

A DNA sequence (GBSx1299) was identified in *S.agalactiae* <SEQ ID 3795> which encodes the amino acid sequence <SEQ ID 3796>. This protein is predicted to be RNA polymerase sigma factor, sigma-70 family (rpoD). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3157(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to the sigma-42 protein from *S.mutans*:

>GP:BAA21507 GB:AB000631 sigma 42 protein [Streptococcus mutans]
 Identities = 345/367 (94%), Positives = 358/367 (97%)

Query: 14 EKKGNTTFNVQVADFIRNHKKQGTATDDEVTETKLVIPFVLADQIDDLERLTDGGISIT 73
 +KK ++TFNVQVADFIRNHKK+G A+DDEVTETKLVIPF L+A+QIDDLERLTDGGISIT
 Sbjct: 5 KKKTSSTFNVQVADFIRNHKKEGVAVDDEVTETKLVIPFELEAEQIDDLERLTDGGISIT 64

Query: 74 DKEGNPSTKYVVEGPKPEELTDEELIGSNSAKVNDPVRMYLKEIGVVPLLTNEEEKELAV 133
 D+EGNPSTKY VE KPEELTDEEL+GSNSAKVNDPVRMYLKEIGVVPLLTNEEEKELA+
 Sbjct: 65 DREGNPSTKYAVEEIKPEELTDEELLGSNSAKVNDPVRMYLKEIGVVPLLTNEEEKELAI 124

Query: 134 AVAEGDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKGFK 193
 AV GDL AKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKGFK
 Sbjct: 125 AVENGDLKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKGFK 184

Query: 194 FSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER 253
 FSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER
 Sbjct: 185 FSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER 244

Query: 254 MDMPDPKVRILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQLDE 313
 MDMPDPKVRILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQLDE
 Sbjct: 245 MDMPDPKVRILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQLDE 304

Query: 314 VLDTLTDREENVLRRLRFLDGDGKMTLEDVGVKVFNVTRERIRQIEAKALRKL RHPSRSKQ 373
 VLDTLTDREENVLRRLRFLDGDGKMTLEDVGVKVF+VTRERIRQIEAKALRKL RHPSRSKQ
 Sbjct: 305 VLDTLTDREENVLRRLRFLDGDGKMTLEDVGVKFDVTRERIRQIEAKALRKL RHPSRSKQ 364

Query: 374 LKDFMED 380
 L+DF+ED
 Sbjct: 365 LRDFVED 371

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3797> which encodes the amino acid sequence <SEQ ID 3798>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

-1371-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1788(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 351/369 (95%), Positives = 364/369 (98%)

10

Query: 12 MAEKKGNITTFNVQVADFIRNHKKQGTAIIDDEVTEKLVIPFVLADADQIDDLLERLTDGGIS 71
 M ++K TTFNVQVA+FIR+HKK+GTAIDD+VTEKLVIPF LDADQIDDLLERLTDGGIS
 Sbjct: 1 MTKQKEITTFNVQVAEFIRHHKKEGTAIDDDVTEKLVIPFALDADQIDDLLERLTDGGIS 60

15

Query: 72 ITDKEGNPSTKYVVEGPKPEELTDEELIGSNSAKVNDPVRMYLKEIGVVPLLTNEEEKEL 131
 ITDKEGNPS+KY+VE PKPEELTDEELIGSNSAKVNDPVRMYLKEIGVVPLLT+EEEEKEL
 Sbjct: 61 ITDKEGNPSSKYIVEEPKPEELTDEELIGSNSAKVNDPVRMYLKEIGVVPLLTSEEEKEL 120

20

Query: 132 AVAVAEGDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKG 191
 AVAVA+GDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKG
 Sbjct: 121 AVAVAKGDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKG 180

25

Query: 192 FKFSTYATWWIRQAITRAIADQARTIRIPVHVMETINKLVREQRNLLQELGQDPTPEQIA 251
 FKFSTYATWWIRQAITRAIADQARTIRIPVHVMETINKLVREQRNLLQELGQDPTPEQIA
 Sbjct: 181 FKFSTYATWWIRQAITRAIADQARTIRIPVHVMETINKLVREQRNLLQELGQDPTPEQIA 240

30

Query: 252 ERM+MTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQ 311
 ERM+MTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQ
 Sbjct: 241 ERM+MTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQ 300

35

Query: 312 DEVLDTLTDREENVLRRLRFGGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKL RHPSRS 371
 DEVLDTLTDREENVLRRLRFGGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKL RHPSRS
 Sbjct: 301 DEVLDTLTDREENVLRRLRFGGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKL RHPSRS 360

Query: 372 KQLKDFMED 380
 KQL+DF+ED

Sbjct: 361 KQLRDFIED 369

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1224

40 A DNA sequence (GBSx1300) was identified in *S.agalactiae* <SEQ ID 3799> which encodes the amino acid sequence <SEQ ID 3800>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2853(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1372-

Example 1225

A DNA sequence (GBSx1301) was identified in *S.agalactiae* <SEQ ID 3801> which encodes the amino acid sequence <SEQ ID 3802>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2198(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA03516 GB:D14690 DNA primase [Lactococcus lactis]

Identities = 206/398 (51%), Positives = 294/398 (73%), Gaps = 6/398 (1%)

Query: 37 LAIDKEKISEIKNSVNIIVDVIGEVVGLTKTGRNHLGLCPFHKEKTPSFNVIEDRQFFHCF 96

+++D E ++++K+ VNI D+I + V L++TG+N++GLCPFH EKTSPFNV ++ F+HCF

Sbjct: 2 VSLDTEVVNDLKSQVNIADLISQYVALSRTGKNYIGLCPFHGEKTPSFNVNAEKGFYHCF 61

Query: 97 GCGRSGDVFKFVEDYQHISFLDSVQVLAERSGIPLDTNFKGQVPPKPKANQSLLDIHRVA 156

GCGRSGD +F+++Y + F+D+V+ LA+ +G+ L N +K N L +I+ A

Sbjct: 62 GCGRSGDAIEFLKEYNQVGFVDAVKELADFAGVTL--NISDDREEKNPNAPLFEINNQA 119

Query: 157 SGFYHAYLMTTNDGERARQYLAERGVTEDLIKHFQIGLSPGGQDFLYRRLAKEFDEKTL 216

+ Y+ LM+T GERAR+YL ERG+T+D+IK F IGL+P DF+++ L+ +FDE+ +

Sbjct: 120 ARLYNILLMSTELGERARKYLEERGITDDVIKRFNIGLAPEENDFIFKNLSNKFDEEIMA 179

Query: 217 SSGLFNYSENSNQFYDSFNRMIFPLTNDIGEVIASFGRVWTQEDIDRKQAKYKNSRATP 276

SGLF++S +N+ +D+F NRIMFP+TN+ G+ I FSGR W QE+ D K AKY N+ AT

Sbjct: 180 KSGLFHFS--NNKVFDAFTNRIMFPITNEYGQTIGFSGRKW-QENDDSK-AKYINTSATT 235

Query: 277 IFNKSYELYHLDKARAVINKAHEVYLMEGFMDVIAAYRAGIENVVASMGTALTNEHVRL 336

IF+KSYEL++LDKA+ I+K HEVYLMEGFMDVIA+Y+AGI NVVASMGTALT +HVR L

Sbjct: 236 IFDKSYELWNLDKAKPTISKQHEVYLMEGFMDVIAASYKAGINNVMVASMGTALTEKHVRL 295

Query: 337 KRFTKKVVLTYDGDRAQNAIDKSLELLSDMTVDIVRIPNKMPPDEFLOANSAEDFKQLL 396

K+ KK VL YDGD AGQNAI K+++L+ + V IV++P +DPDE+ + + L+

Sbjct: 296 KQMAKKFVLVYDGDAGQNAIYKAIDLIGESAVQIVKVPGLDPDEYSKNYGLKGLSALM 355

Query: 397 ENGRISNTEFYIHYLKPENTDNLQSEIAYVEKIAKLI 434

E GRI EF I YL+PEN NLQ+++ ++E+I+ +IA

Sbjct: 356 ETGRIQPIEFLLIDYLRPENLANLQTLDFIEQISPMIA 393

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3803> which encodes the amino acid sequence <SEQ ID 3804>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3532(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 378/604 (62%), Positives = 477/604 (78%), Gaps = 2/604 (0%)

Query: 28 MGYFCGGHDLAIDKEKISEIKNSVNIIVDVIGEVVGLTKTGRNHLGLCPFHKEKTPSFNVI 87

MG+ GG DLAIKKE IS++KNSVNIIVDVIGEVV L+++GR++LGLCPFHKEKTPSFNV+

Sbjct: 1 MGFLWGGDDLAIKEMISQVKNSVNIIVDVIGEVVKLRSRGRHYLGLCPFHKEKTPSFNVV 60

Query: 88 EDRQFFHCFGCGRSGDVFKFVEDYQHISFLDSVQVLAERSGIPLDTNFKGQV--PKKPKA 145

-1373-

EDRQFFHCFGCG+SGDVFKF+E+Y+ + FL+SVQ++A+++G+ L+ V +
 Sbjct: 61 EDRQFFHCFGCGKSGDVFKFIEEYRQVPFFLESVQIIADKTGMSLNIPPSQAVLASQHKHP 120

Query: 146 NQSLLDIHRVASGFGYHAYLMTTNDGERARQVIAERGVTEDLIKHFQIGLSPGGQDFLYRR 205
 N +L+ +H A+ FYHA LMTT G+ AR+YL +RG+ + LI+HF IGL+P D+LY+
 Sbjct: 121 NHALMTLHEDAAKFYHAYLMTTITIGQEARLYQRLDDQLIEHFNIGLAPDES DYLYQA 180

Query: 206 LAKEFDEKTLMSGLFNYSSENSNQFYDSFNNRIMFPLTNDIGEVIASFGRVWTQEDIDRK 265
 L+K+++E L++SGLF+ S+ SN YD+F NRIMFPL++D G +IAFSGR+WT D++++
 Sbjct: 181 LSKKYEGLVAVSGLFHLSDQNTIYDAFRNRIMFPLSDDRGHIIAFSGRIWTAADMEKR 240

Query: 266 QAKYKNSRATPIFNKSYELYHLDKARAVINKAHEVYLMEGFMDVIAAYRAGIENVVASM 325
 QAKYKNSR T +FNKSYELYHLDKAR VI K HEV+LMEGFMDVIAAYR+G EN VASM
 Sbjct: 241 QAKYKNSRGTVLFNKSYELYHLDKARVIAKTHEVFLMEGFMDVIAAYRSGYENAVASM 300

Query: 326 TALTNEHVRHLKRFTHKKVVLTYDGDAGQNAIDKSLELLSDMTVDIVRI PNKMDPDEFLQ 385
 TALT EHV HLK+ TKKVVL YDGD AGQ+AI KSLELL D V+IVRI PNKMDPDEF+Q
 Sbjct: 301 TALTQEHVNHKQVTKKVVLTYDGDAGQHAIAKSLELLKDFVVEIVRI PNKMDPDEFVQ 360

Query: 386 ANSAEDFKQLLENGRISNTEFYIHYLKPENTDNLQSEIAYVEKIAKLIKSPSITAQNSY 445
 +S E F LL+ RIS+ EF+I YLKP N DNLSQ+I YVEK+A LIA+SPSITAQ+SY
 Sbjct: 361 RHPSEAFADLLKQSRISSEVEFFIDYKPTNVDNLQSQIVYVEKMAPLIAQSPSITAQHSY 420

Query: 446 ITKVAELLPDFDYFQVEQSVNNERLHRSQQQASSSVQTSATVQLPQTGKLSAITKTEMQ 505
 I K+A+LLP+FDYFQVEQSVN R+ R + Q + S V LP L+AI KTE
 Sbjct: 421 INKIADLLPNFDYFQVEQSVNALRIQDRQKHQGIQAQAVSNLVTLPMPKSLTAIAKTESH 480

Query: 506 LFHRLLNHPYLLNEFRNRDNFYFDTTTEIQVLYELLKESGEITSYDLSQESDKVNRTYYII 565
 L HRL+H YLLNEFR+RD+FYFDT+ +++LY+ LK+ G ITSYDLS+ S++VNR YY +
 Sbjct: 481 LMHRLHLHDYLLNEFRNRDDFYFDTSTLELLYQRLKQGHITSYDLSSEMSEEVNRAYYNV 540

Query: 566 LEEQLPVEVSIGIEIAVEKARDRLKQSQLIRQSSNQDEEGALAALENLIAQK 625
 LEE LP EV++GEI+ + R +LL ERDL KQ + +R+SSN+GD + AL LE+ IAQK
 Sbjct: 541 LEENLPKEVALGEIDDILSKRAKLLAERDLHKQKGVRESSNKGDHQAALVLEHFIAQK 600

Query: 626 RNME 629
 R ME
 Sbjct: 601 RKME 604

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1226

A DNA sequence (GBSx1302) was identified in *S. agalactiae* <SEQ ID 3805> which encodes the amino acid sequence <SEQ ID 3806>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.05 Transmembrane 41 - 57 (34 - 58)
 INTEGRAL Likelihood = -5.79 Transmembrane 93 - 109 (90 - 112)

----- Final Results -----
 bacterial membrane --- Certainty=0.3421(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9995> which encodes amino acid sequence <SEQ ID 9996> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC38560 GB:AF029731 large conductance mechanosensitive channel
 [Staphylococcus aureus]
 Identities = 64/126 (50%), Positives = 83/126 (65%), Gaps = 8/126 (6%)

-1374-

Query: 23 MIKELKEFLFKGNVLDLAVAVILGAAFNAIITSLVKDVITPLILNPVLKAAGVSNIA-QL 81
 M+KE KEF KGNVLDLA+AV++GAAFN II+SLV+++I PLI K G + A +
 5 Sbjct: 1 MLKEFKEFALKGNVLDLAIAVVMGAAFNKIISSLVENIIMPLI----GKIFGSVDFAKEW 56

Query: 82 SWNGVAYGNFLSAVINFLIVGTTLFFIVKAANKVMAKKPAEEEEIEVVEPTQEQLLAEIR 141
 S+ G+ YG F+ +VI+F+I+ LF VK AN +M K+ AEE E V LL EIR
 Sbjct: 57 SFWGIKYGLFIQSVIDFTIIAFALFIFVKIANTLMKKEEAE---EAVVEENVVLLTEIR 113

10 Query: 142 DLLANK 147
 DLL K
 Sbjct: 114 DLLREK 119

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3807> which encodes the amino acid
 15 sequence <SEQ ID 3808>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -5.95 Transmembrane 71 - 87 (67 - 90)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.3378(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:CAB15653 GB:Z99122 similar to large conductance mechanosensitive
 channel protein [Bacillus subtilis]
 Identities = 61/126 (48%), Positives = 77/126 (60%), Gaps = 7/126 (5%)

30 Query: 1 MVKELKAFLFRGNIIE LAVAVIIGGAFGAIVTSFVNDIITPLILNPALKAAVENITQLS 60
 M E KAF RGNI++LA+ V+IGGAFG IVTS VNDII PL+ L + ++
 Sbjct: 1 MWNEFKAFAMRGNIVDLAIGVVIGGAFGKIVTSLVNDIIMPLV-GLLLGGLDFSGLSFTF 59

35 Query: 61 WNG-VKYGSFLGAVINFLIIGTSLFFVVKAAEKAMPKE-----KEAAAPTQEELLTEIR 114
 + VKYGSF+ ++NFLII S+F V++ KKE E A QEELL EIR
 Sbjct: 60 GDAVVKYGSFIQTIVNFLIISFSIFIVIRTLNGLRRKKEAEAEAEAEVDAQEELLKEIR 119

Query: 115 DLLAQK 120
 DLL Q+
 40 Sbjct: 120 DLLKQQ 125

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/125 (68%), Positives = 99/125 (78%), Gaps = 5/125 (4%)

45 Query: 23 MIKELKEFLFKGNVLDLAVAVILGAAFNAIITSLVKDVITPLILNPVLKAAGVSNIAQLS 82
 M+KELK FLF+GN+++LAVAVI+G AF AI+TS V D+ITPLILNP LKAA V NI QLS
 Sbjct: 1 MVKELKAFLFRGNIIE LAVAVIIGGAFGAIVTSFVNDIITPLILNPALKAAVENITQLS 60

50 Query: 83 WNGVAYGNFLSAVINFLIVGTTLFFIVKAANKVMAKKPAEEEEIEVVEPTQEQLLAEIRD 142
 WNGV YG+FL AVINFLI+GT+LFF+VKAA K M KK E PTQE+LL EIRD
 Sbjct: 61 WNGVKYGSFLGAVINFLIIGTSLFFVVKAAEKAMPKKEK-----EAAAPTQEELLTEIRD 115

Query: 143 LLANK 147
 LLA K
 55 Sbjct: 116 LLAQK 120

A related GBS gene <SEQ ID 8753> and protein <SEQ ID 8754> were also identified. Analysis of this
 protein sequence reveals the following:

60 Lipop: Possible site: -1 Crend: 10
 SRCFLG: 0
 McG: Length of UR: 4
 Peak Value of UR: 2.96

-1376-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9391> which encodes amino acid sequence <SEQ ID 9392> was also identified. A related GBS nucleic acid sequence <SEQ ID 10799> which encodes amino acid sequence <SEQ ID 10800> was also identified.

The protein is similar to the 30S ribosomal protein S21 from *Listeria monocytogenes*:

>GP:BAA82793 GB:AB023064 30S ribosomal protein S21 [*Listeria monocytogenes*]
Identities = 30/34 (88%), Positives = 34/34 (99%)

Query: 1 MTKAGTLQESRKREFYEKPSVKRKRKSEAARKRK 34
++K+GTLQESRKREFYEKPSVKRKRKSEAARKRK
Sbjct: 23 VSKSGTLQESRKREFYEKPSVKRKRKSEAARKRK 56

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3811> which encodes the amino acid sequence <SEQ ID 3812>. Analysis of this protein sequence reveals the following:

Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4815(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 35/36 (97%), Positives = 36/36 (99%)

Query: 1 MTKAGTLQESRKREFYEKPSVKRKRKSEAARKRK 36
+TKAGTLQESRKREFYEKPSVKRKRKSEAARKRK 36
Sbjct: 35 VTKAGTLQESRKREFYEKPSVKRKRKSEAARKRK 70

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1228

A DNA sequence (GBSx1304) was identified in *S.agalactiae* <SEQ ID 3813> which encodes the amino acid sequence <SEQ ID 3814>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -7.06 Transmembrane 5 - 21 (3 - 23)
INTEGRAL Likelihood = -2.28 Transmembrane 191 - 207 (189 - 207)

----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8755> and protein <SEQ ID 8756> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 2
McG: Discrim Score: 8.68
GvH: Signal Score (-7.5): -5.71

-1377-

Possible site: 18
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 2 value: -7.06 threshold: 0.0
 INTEGRAL Likelihood = -7.06 Transmembrane 5 - 21 (3 - 23)
 INTEGRAL Likelihood = -2.28 Transmembrane 191 - 207 (189 - 207)
 PERIPHERAL Likelihood = 4.35 142
 modified ALOM score: 1.91

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8756 (GBS259) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 4; MW 54kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1229

A DNA sequence (GBSx1305) was identified in *S.agalactiae* <SEQ ID 3815> which encodes the amino acid sequence <SEQ ID 3816>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 136 - 152 (135 - 152)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD47593 GB:AF140784 Vexp2 [Streptococcus pneumoniae]
 Identities = 117/212 (55%), Positives = 152/212 (71%)
 Query: 1 MLELKNIAYRYKGNDNKTLENINYSFQSGVFYITLGNSSGSKTLLSLMAGLDSPTGEQV 60
 +L+L+++ YRYK L INY+F+ G FY+I+G SG+GK+TLLSL+AGLDSP EG +
 Sbjct: 3 LLQLQDVITYRYKNTAEAVLYQINYNFEPGKFYSIIGESGAGKSTLLSLLAGLDSPVEGSI 62
 Query: 61 LFNKKDIKEAGYAQHRKKNIALVFQNYNLLDYLTPLENVQLVKPTADKQLLLDLGLKEDM 120
 LF +DI++ GY+ HR +I+LVFQNYNL+DYL+PLEN++LV A K LL+LGL E
 Sbjct: 63 LFQGEDIRKKGYSYHRMHHSILVFQNYNLLIDYLSPLENIRLVNKKASKNTLLELGLDESQ 122
 Query: 121 LTRNLRSLSGGQQQRVAIARALVVGTPAILLDEPTGNLDFDISRDITMRLKDFAHKEKRC 180
 + RN+L+LSGGQQQRVAIAR+LV P IL DEPTGNLD + DI LK A K +C
 Sbjct: 123 IKRNVQLSLGGQQQRVAIARSLVSEAPVILADEPTGNLDPRTAGDVELLKSQAQKTGKC 182
 Query: 181 VIMVTHSREIAHMADTALQLIGDNLKELSKES 212
 VI+VTHS+E+A +D L+L L E S
 Sbjct: 183 VIVVTHSKEVAQASDITLELKDKKLTETRNIS 214

SEQ ID 3816 (GBS363) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 5; MW 28kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 10; MW 53kDa).

GBS363-GST was purified as shown in Figure 216, lane 9.

-1378-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1230

A DNA sequence (GBSx1306) was identified in *S.agalactiae* <SEQ ID 3817> which encodes the amino acid sequence <SEQ ID 3818>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have an uncleavable N-term signal seq
10  INTEGRAL    Likelihood = -14.97    Transmembrane    71 - 87 ( 66 - 97)
    INTEGRAL    Likelihood = -3.61    Transmembrane     2 - 18 ( 1 - 18)

----- Final Results -----
    bacterial membrane --- Certainty=0.6986(Affirmative) < succ>
15    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1231

A DNA sequence (GBSx1307) was identified in *S.agalactiae* <SEQ ID 3819> which encodes the amino acid sequence <SEQ ID 3820>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

25  Possible site: 45
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1986(Affirmative) < succ>
30    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1232

A DNA sequence (GBSx1308) was identified in *S.agalactiae* <SEQ ID 3821> which encodes the amino acid sequence <SEQ ID 3822>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

45  Possible site: 34
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -6.05    Transmembrane    22 - 38 ( 17 - 39)

----- Final Results -----
    bacterial membrane --- Certainty=0.3421(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```


-1379-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAD47594 GB:AF140784 Vexp3 [Streptococcus pneumoniae]
Identities = 39/153 (25%), Positives = 67/153 (43%), Gaps = 9/153 (5%)

Query: 3 LFKRSFLYVSRKKRSITL FVCLWLVASTLISGIAVKNAGLTA-KKTFSRQTGSILHISS 61
+ +F YV+RK KSI +F+ + L+AS + G+++K A A ++TF T S +
10 Sbjct: 1 MLHNAFAYVTRKFFKSIVIFLIILLMASLSLVGLSIKATAKASQETFKNITNS-FSMQI 59

Query: 62 DSTDLVGDGYGSGEIPKAI VNIASNPVNRVNNLMAYAGLTSEKMTVRPNDKEQYKE- 120
+ G G+G I + I I N ++ + A LT ++ P K+
Sbjct: 60 NRRVNQGTTPRGAGNIKGEDIKKITEKAI ESYVVRINAIGDLTGYDLIETPETKKNLTD 119

15 Query: 121 -----QVLQVHGNSYSDDPKYTAGMISLKGK 147
L + G + S + K+ +G L G
Sbjct: 120 RAKRFGSSLMITGVNDSSKEDK FVSGSYKLVEG 152

No corresponding DNA sequence was identified in *S.pyogenes*.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1233

A DNA sequence (GBSx1309) was identified in *S.agalactiae* <SEQ ID 3823> which encodes the amino acid sequence <SEQ ID 3824>. Analysis of this protein sequence reveals the following:

25 Possible site: 39
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -15.76	Transmembrane	295 - 311 (287 - 317)
INTEGRAL	Likelihood = -7.59	Transmembrane	49 - 65 (46 - 69)
INTEGRAL	Likelihood = -6.90	Transmembrane	340 - 356 (339 - 362)
30 INTEGRAL	Likelihood = -5.57	Transmembrane	411 - 427 (404 - 430)

----- Final Results -----

	bacterial membrane --- Certainty=0.7305(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9695> which encodes amino acid sequence <SEQ ID 9696> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB12182 GB:Z99106 similar to transporter [Bacillus subtilis]
Identities = 95/370 (25%), Positives = 167/370 (44%), Gaps = 41/370 (11%)

Query: 109 ESVEASLSIDVGSRLKSVSPYNSS-----KEENQVTLAGYQSTEDLRAFQTKALVLK 160
+++E+S S D S S + NS + +++ G ST + F +
45 Sbjct: 115 DAIESSSSSSSSSSSSSNAKNSQGGGQGGPQMVQADLSIEGVISTALVDDFSDGDSKIT 174

Query: 161 KGSHLAADNT--KQVLVPLKLAQKNHLSVGNKLRGK---ENVT----IAGIYDANSA-- 209
G + + K ++ LA++N LSVG+ ++ E+ T I GIY S+
Sbjct: 175 DGRAITKSDVGKKVTVINETLAEENDLSVGDSITTESATDEDTTVKLKIVGIYKTTSSGD 234

50 Query: 210 -KSKNTFNPNIDNTLIAQATLVKRLSKQKGYQTV---AVRLSDKRLVDTVIQNIKQWPLD 265
+++N N N L T + T+ + D + +DT ++ K+ +D
Sbjct: 235 DQAQNFSLNPYNKLYTPYTATAALKGDDYKNTIDS AVYYMDDAKNMDTFVKAAKKTSID 294

55 Query: 266 FGKLDVQTAKEFYGDSYRNIE TLHRLVGRILIVSLVAMAILVVMILTFWINNRIKETGIL 325
F + T + Y IE + ++ +VS+ IL +++ I R E G+L
Sbjct: 295 FDTYTLNTNDQLYQQMVGP IENVASF SKNVVYLVS VAGAVILGLIVMMSIRERKYEMGVL 354

5
10

SQGGGQGGPQMVQADLSIEGVISTALVDDFSDGDSKITDGRAITKSDVGKKVTVINETLAEENL

-1381-

```

      885      903      954      978      1008      1065
      NKRLRL---GKENVTI---AGIYDANSA---KSKNTFNPNDNTLIA--QATLVRKISKQKGYQTVAVR--LSDKRLVDTV
      : : : | : | : || | : : : | | | | | | | | : | : : |
5      DSITIESATDEDTTVKLKIVGYKTTSSGDDQAQNFSLNPNKLYTPYTATAALKGDDYKNTIDSAVYYMDDAKNMDTF
      220      230      240      250      260      270      280

      1095      1125      1155      1185      1215      1245      1275      1305
      IQNIKQWPLDFGKLDVQTAKEFYGDSYRNIE TLHRLVGR IILIVSLVAMAILVVMLTFWVNNRIKETGILLAIGKTKFEI
      : : | : : | | : | : : | | : : : | : | : : | : : : : :
10      VKAAKTSIDFDITYTLNNDQLYQQMVGP IENVASF SKNVVLVSVAGAVILGLIVMMSIRERKYEMGVLMAIGEKRWKL
      300      310      320      330      340      350      360

      1335      1365      1395      1431      1461      1491
      IGHYLIEVLLVAGAAFTLSIIGVFLGKTFAAGLLSQV-----NGGVSSQIVQNSSLIIDRIDNLAV
      || : | : | : | | : : | : : | | | | | : : : | : | : | : |
15      IGQFLTEILIVAVIAIGLASVTGNLVANQLGNQLLSQQIISSTDSTQTASGQMPGGGGMGGMFGHSSSNVDVIDSLNV
      380      390      400      410      420      430      440

      1521      1551      1581      1611      1641      1671      1701      1731
      SVGVMDVFRLYAQGALICLFAVLSSYSILKLPKQILSRMS*EVNMNLFKRSFLYVSRKKRSITLFLVCLWLVLASTLIS
      : | : : | | | : | : | | : | : | : | : :
20      AVSMNDMLILGGIGILIAIATLLPSISVLRHLHPKTI LTKQE
      460      470      480

```

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1234

A DNA sequence (GBSx1310) was identified in *S.agalactiae* <SEQ ID 3825> which encodes the amino acid sequence <SEQ ID 3826>. Analysis of this protein sequence reveals the following:

```

30      Possible site: 24
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----
35      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40      >GP:CAB11993 GB:Z99105 ybdG [Bacillus subtilis]
      Identities = 66/224 (29%), Positives = 102/224 (45%), Gaps = 22/224 (9%)

      Query: 84  IKEYGQKVEVKGKKMNVYTVGEGKVPIVFIPGQGTVTAKHQYHNLSNLSTHKVTVVVEP 143
      +K G V+V GKMMNVY G GK VF+ G G ++ L S SK +K+ VV+
45      Sbjct: 41  LKGKGTVDVDGKKMNVYQEGSGKDTFVFMMSGGIAAPAYEMKGLYSKFSENKIAVVDR 100

      Query: 144  FGSLSDVIDQPRNLANITSDIHEALQKVGITGKYVIASHSIGGVYALKYISTYPKEVLG 203
      G G S+V R++ + +AL K G Y++ HSI G+ A+ + YPKE+
      Sbjct: 101  AGYGYSEVSHDDRDIDTVLEQTRKALMKSGNKPPYILMPHSISGIEAMYWAQKYPKEIKA 160

50      Query: 204  LIGLDTSTP-----GMEGGKQVDF-----AAPVLKELPKIPKVSDDIN 241
      +I +D P G++ K F +A E+ + ++D+
      Sbjct: 161  IIAMDIGLPQQYVVTYKLSGVDRLKVRGFHLLTSIGFHRFIPS AVYNPEVIRQSFLTDEEK 220

      Query: 242  AQFFAIGHKILNNSNMKEEAKNSSNMINESANYKIPKGIPAMYL 285
      + AI K N++M+ E S ++S N PK P + L
55      Sbjct: 221  EIYKAINFKQFFNADMEHELLQSYQNGSKSVNLPAPKETPVIL 264

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-1382-

SEQ ID 3826 (GBS121) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 9; MW 40kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 6; MW 65kDa).

GBS121-GST was purified as shown in Figure 198, lane 6.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1235

A DNA sequence (GBSx1311) was identified in *S.agalactiae* <SEQ ID 3827> which encodes the amino acid sequence <SEQ ID 3828>. Analysis of this protein sequence reveals the following:

```

10   Possible site: 33
    >>> Seems to have a cleavable N-term signal seq.

    ----- Final Results -----
    bacterial outside --- Certainty=0.3000(Affirmative) < succ>
15   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8759> which encodes amino acid sequence <SEQ ID 8760> was also identified. Analysis of this protein sequence reveals the following:

```

20   Lipop: Possible site: -1   Crend: 8
    McG: Discrim Score:      3.70
    GvH: Signal Score (-7.5): -0.0600004
    Possible site: 22
    >>> Seems to have a cleavable N-term signal seq.
25   ALOM program  count: 0 value:  8.01 threshold:  0.0
    PERIPHERAL Likelihood =  8.01      167
    modified ALOM score: -2.10

    *** Reasoning Step: 3
30   ----- Final Results -----
    bacterial outside --- Certainty=0.3000(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35   bacterial cytoplasm --- Certainty=0.0000(Not Clear)

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8760 (GBS60) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 7; MW 38.6kDa).

GBS60-His was purified as shown in Figure 193, lane 3.

- 40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1236

A DNA sequence (GBSx1312) was identified in *S.agalactiae* <SEQ ID 3829> which encodes the amino acid sequence <SEQ ID 3830>. This protein is predicted to be unnamed protein product. Analysis of this

- 45 protein sequence reveals the following:

```

    Possible site: 21
    >>> May be a lipoprotein

```


-1384-

GBS21L was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 124 (lane 8-10; MW 66.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 124 (lane 11; MW 41.5kDa) and in Figure 180 (lane 6; MW 41kDa). GBS21L-His was purified as shown in Figure 232 (lanes 3 & 4)

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1237

A DNA sequence (GBSx1313) was identified in *S.agalactiae* <SEQ ID 3831> which encodes the amino acid sequence <SEQ ID 3832>. This protein is predicted to be endopeptidase O. Analysis of this protein
10 sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3854(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAF67832 GB:AF179267 endopeptidase PepO2 [Lactococcus lactis]
Identities = 21/36 (58%), Positives = 26/36 (71%)

Query: 1 MRANIPVRNFQEFYDAFGVKKGDSEMYLKPEKRLTLW 36
+RANIP N +EFY+ F VK+ D MY PEKRL +W

25 Sbjct: 592 LRANIPPTNLEEFYETFDVKETDQMYRAPEKRLKIW 627

There is also some homology to SEQ ID 2384:

Identities = 13/36 (36%), Positives = 25/36 (69%)

30 Query: 1 MRANIPVRNFQEFYDAFGVKKGDSEMYLKPEKRLTLW 36
+R N+ + NF F++ F +K+GD+M+ P+ R+ +W
Sbjct: 596 LRTNVTLTNFDAPHETFDIKEGDAMWRAPKDRVIIW 631

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1238

A DNA sequence (GBSx1314) was identified in *S.agalactiae* <SEQ ID 3833> which encodes the amino acid sequence <SEQ ID 3834>. This protein is predicted to be endopeptidase O. Analysis of this protein sequence reveals the following:

40 Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3801(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAA16168 GB:L18760 endopeptidase [Lactococcus lactis]
Identities = 118/268 (44%), Positives = 174/268 (64%), Gaps = 6/268 (2%)

-1385-

Query: 1 MGDYVGKKYFGEAAKDV EHM AKKI INVYKTR LKNNTWLS ENT KAMA I KKL DNMRL MIGY 60
 +G +Y GKKYFGEAAK DV+ M +I VY+ RL N WLS+ T AI+KLD + IG+
 Sbjct: 321 IGLFYGKKYFGEAAKADV KRMV TAMI KVVQVRLSKNEWLSQETA EKAEKLDAITPFIGF 380

Query: 61 PEDYPDLYRQYQFDSKASFFENNDNYR KLSNKKTFEEFNQSNQREHWQMSANAVNAYNDP 120
 P+ P++Y + + S S +E+ + K+ +TFE+F++ + W M A+ VNAY P
 Sbjct: 381 PDKLPEIYSRLKTT S-GSLYEDALKF D KILTARTFEKFS EDVDKTSWHMPAHMVNAYYSP 439

Query: 121 NTNSIVFPAAIFQSPLYDKTKTVSQNYGAIGA IIGHEISHSFDINGMKYDEKGNLHDWWT 180
 ++N+IVFPAAI Q+P Y ++ SQNYG IGA+I HEISH+FD NG ++D++GNL+ WW
 Sbjct: 440 DSNITIVFPAAILQAPFYSL EQSSSQNYGGIGAVIAHEISHAFDNNGAQFDKEGNLKNKWWL 499

Query: 181 KEDLKHYKKKTQAMIDQWDGLKADGGKVDGKLT LAENIADNGGVMASLEALKTEKIQTIK 240
 ED + +++K + MI +DG++ + G +GKL ++ENIAD GG+ A+L A K EK +K
 Sbjct: 500 DEDYEAFEEKQKEMIALFDGVETEAGPANGK LIVSENIADQGGITAALTA AKDEKDVDLK 559

Query: 241 NFLNHGQVFGVKKQPKNKVSPQFSQMFM 268
 F + K + K S +F QM +
 Sbjct: 560 AFFSQW-----AKIWRMKASKEFQQMLL 582

There is also homology to SEQ ID 2384:

Identities = 110/253 (43%), Positives = 161/253 (63%), Gaps = 1/253 (0%)

Query: 1 MGDYVGKKYFGEAAKDV EHM AKKI INVYKTR LKNNTWLS ENT KAMA I KKL DNMRL MIGY 60
 +G +Y + F AK DVE ++I VYK+RL+ WL+ T+ AI KL+ + IGY
 Sbjct: 324 LGLWYAGQKFSPEAKADVESKVAR MIEVYKSRL ETADWLAPATREKAITKLN VITPHIGY 383

Query: 61 PEDYPDLYRQYQFDSKASFFENNDNYR KLSNKKTFEEFNQSNQREHWQMSANAVNAYNDP 120
 PE P+ Y + D S EN N K++ T+ ++N+ R W M A+ VNAY D
 Sbjct: 384 PEKLPETYAKKVIDESLSLIVENAQNLAKITIAHTWSKWNKPVD RSEWHMPAHLVNAYYDL 443

Query: 121 NTNSIVFPAAIFQSPLYDKTKTVSQNYGAIGA IIGHEISHSFDINGMKYDEKGNLHDWWT 180
 N IVFPAAI Q P Y ++ S NYG IGA+I HEISH+FD NG +DE G+L+DWWT
 Sbjct: 444 QONQIVFPAAILQEPFYSLDQSSSANYGGIGAVIAHEISHAFDINGASFDEHGS LNDWWT 503

Query: 181 KEDLKHYKKKTQAMIDQWDGLKADGGKVDGKLT LAENIADNGGVMASLEALKTEKIQTIK 240
 +ED +K++T ++ Q+DGL++ G KV+GKLT++EN+AD GGV +LEA ++E+ + +
 Sbjct: 504 QEDYAAFKERTDKIVAQFDGLES HGAKVNGKLT VSENVADLGGVACALEAAQSEEDFSAR 563

Query: 241 N-FLNHGQVFGVK 252
 + F+N ++ +K
 Sbjct: 564 DFFINFATIWRMK 576

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1239

A DNA sequence (GBSx1315) was identified in *S.agalactiae* <SEQ ID 3835> which encodes the amino acid sequence <SEQ ID 3836>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9691> which encodes amino acid sequence <SEQ ID 9692> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1386-

>GP:AAC35997 GB:AF019410 endopeptidase O [*Lactobacillus helveticus*]
Identities = 85/315 (26%), Positives = 146/315 (45%), Gaps = 8/315 (2%)

Query: 46 NVSPRENLYRAVNDNLANTKLKQGQTSVNSFSEIEDKLKQLLVSDMAKMASGKIETTN- 104
N P++NLY AVN WL+ ++ QTS +E++ K+++ ++ D A +ASGK + +
Sbjct: 20 NAKPQDNLYLAVNSEWLSKAEPADQTSAGVNTELDIKIEKRMMKDFADIASGKEKMPDI 79

Query: 105 DEQKKMVAYYKQGMDFKTRDKNGLKPLKPVLOKLEAVSSMKDFQSLAHDFVMSGFVLPPF 164
+ K +A YK +F RD P++ LQK+ + + F+ A + M + LPF
Sbjct: 80 RDFDKAIALYKIAKNFDRDAEKANPIQNDLQKILDILINFDKFDNATELFMGYPALPFV 139

Query: 165 LTVETNARDNSQKQLVLRQAPALLESFQYKKGNGEKEAKLSAYRTSAMALLKQAGKSNI 224
V+ + ++ L L YK E + L ++ LL+ AG
Sbjct: 140 FDVDADMKNTDFNVLFHFGGPSTFLPDTTTTYK--TPEAKLLDILEKQSINLLEMAGIGKE 197

Query: 225 EDRKLVKQAIADFRLLEKTSQVDSKITAESETAAGRYNPESMETVHNYAKEFDKELIE 284
E R V+ A+AFD+ LS+ K T E A YNP S+ K FD + ++
Sbjct: 198 EARVYVQNALAFDQKLSKVV-----KSTEEWSDYAAIYNPVSLEFLAKFKSFDMAFLK 252

Query: 285 KLVGPTNKAVNVEDKTYFKQVNDVINSKQLANMKAWMMISMLVDQSDFLGEQNRQAASAF 344
++ + V V + + +++IN +K WM++ + + +L + R AA F
Sbjct: 253 TILPEKVERVIVMEPRFLDHADELINPANFDEIKGWMLVKYINSVAKYLSQDFRAAAFPF 312

Query: 345 KNVASGLTQIESKEK 359
SG ++ S+ K
Sbjct: 313 NQAISGTPELPSQIK 327

A related GBS gene <SEQ ID 8763> and protein <SEQ ID 8764> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: 5.41
GvH: Signal Score (-7.5): -1.39
Possible site: 36
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 2.76 threshold: 0.0
PERIPHERAL Likelihood = 2.76 151
modified ALOM score: -1.05

*** Reasoning Step: 3

----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8764 (GBS12) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 7; MW 65kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 3; MW 39kDa).

The GST-fusion protein was purified as shown in Figure 189, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1240

A DNA sequence (GBSx1317) was identified in *S.agalactiae* <SEQ ID 3839> which encodes the amino acid sequence <SEQ ID 3840>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.75 Transmembrane 301 - 317 (299 - 317)

-1387-

----- Final Results -----

bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB42180 GB:A67181 unnamed protein product [unidentified]
 Identities = 245/771 (31%), Positives = 410/771 (52%), Gaps = 80/771 (10%)

10 Query: 22 VRVIVEFNKESILDYATEQKKITVAQLNQADVEKKLQSIKQEQDKVLKNIKESVHFDSSKV 81
 VRVIV NK + D+ ++ + A + + +E+ +K Q+KV+K +E+ +KV
 Sbjct: 97 VRVIVSLNKSAAFDHTSKPTGSAASVKK--IEQASDQVKDQGEKVIKQVEE---ITGNKV 151

15 Query: 82 KR-YDAIINGVALDIQAQIEKLTADVRRVYVSQEVYQTKPLLSSSGQLIGLPEVWNN 140
 +R + ++N ++D+ +I+K+K + V+ V + Y P S+ Q+ + +VW
 Sbjct: 152 RRQFGYLVNAFSIDMDLDDIDKVKDLQVKNVTPVKVY---HPTDESADQMAQVQDVWQE 208

20 Query: 141 SQYKGEGETVVAVIDSGVDFKHQALKIKEPNRAKYNKTSIE----KLIHEKNLKGKFYSEK 196
 + KGEV V+++ID+G+D HQ LK+ +K+ +E KL H GK+Y+EK
 Sbjct: 209 QKLKGEGMVISIIDTGIDSSHQDLKLDGSGVSTALSKSEVESDKSKLGH-----GKYYTEK 263

25 Query: 197 VPYGYNYDYNDNLKDS-YGVMHGMHVTGIVGANDDNQKLYGVAPNAQILAMKVFSDDQQ 255
 VPYGYNY D ND + D+ G MHG HV GI GAN ++ GVAP+AQ+LAMKVFS++ +
 Sbjct: 264 VPYGYNYADKNDQIVDNGCGEMHGQHVAGIAGANG---QVKGVPDAQLLAMKVFSNNAK 320

30 Query: 256 NPTTFTDVWLKALDDAILLKADVNNMSLGTGAGFVHEGKDYPELEVIRACKAGIVIAVA 315
 N + D + A++D++ L ADV+NMSLG+ + V G P+ + +A+A +AG++ ++
 Sbjct: 321 NSGAYDDDIISAIEDSVKLADVINMSLGSVSSDV--GPSDPQQQAVAKASEAGVINVIS 378

35 Query: 316 AGNE---GNITDGNTYGVKPLAENYDTALIANPALDDNTLAVASMENLKKHAHVLKFK-- 370
 AGN G+ DGN +E + + P + + L VAS EN K +K +
 Sbjct: 379 AGNSGVAGSTADGNPNVNTGTSE---LSTVGTGVTPTDALTVASAENSKVTTDTVKDELG 435

40 Query: 371 -----DKKSGTEVTEVINLHVAPNASKTIIGLAVDLGAGAPSELS--KHFDLSGKIA 420
 + K +VT + + + K + VD+G G + + K ++ G++A
 Sbjct: 436 GVTFSNSELKGAAQVTTQLESNYSVLTKLKL---VDMGLGGADDYTAEKKAIEVKGQLA 492

45 Query: 421 MLEIPEDNKSNGFLEKQVQAITKLNPAAILLYNNAKVKDDLGSQLLVESEAAKFNIARITR 480
 +++ + F KV A I++YN+ D L S L + +++
 Sbjct: 493 VVK---RGAYTFSAKVANAKAAGAAGIVYNSE--DDGLLSMSLDDKTFTPLGMSKADG 546

50 Query: 481 STY----NNIKNNSNKIITILTERQAINSLAGQLSSYSSWGPTPDRLKPEITAPGGHI 536
 + ++ + K T L IDNS AG++S ++SWGPTP+L KPEITAPGG I
 Sbjct: 547 KFWLKQKKVRASRLKFGTAL-----IDNSRAGKMSDFTSWGPTPELDFKPEITAPGGKI 601

55 Query: 537 FSTVEDNQYADKSGTSMAPQVAGAAVVKQYITDKKIPV--DNAADFIKLLMNTAQPI 594
 +S DN+Y SGTSMAP VAG+ A++ Q I + + + + F K MNT+ P+
 Sbjct: 602 YSLANDNKYQQMSGTSMASPFVAGSEALILQGIKKQGLNLSGEELVQFAKNSAMNTSHPV 661

60 Query: 595 IN-KQSKDGKTPYFVRQQSGGAMNLAALVTTVVATVTGTNDNNDGKLELREL-KEKKF 652
 + + +K+ +P R+QSGG +N+ A+ TV N +G L+E+ ++ F
 Sbjct: 662 YDTEHTKEIISP---RRQGSGEINVKDAINNTVEVKAA-----NGNGAAALKEIGRQTTF 713

65 Query: 653 KARILLRNFGKTNKTYIISSEA--IADPVDEKGFRTQNSEHLVSKKADAVTRKVTVEAGK 710
 K + L N GK +TY + + + K +++ +V + T KVT+ G+
 Sbjct: 714 K--VTLTNHGKKAQTYAVDNYGGPYTQATEAKSGEIDTK-IVKGQLTTETPKVTVPQGE 770

Query: 711 TLAVDLDDVDYSDAEALTRNNFLEGYLNK-DTEGVADLHLPFLGFYGSWTE 760
 +VD+ + + R NF+EGY+ + + +L LP++GF+GS+++
 Sbjct: 771 --SVDVSFTLTLPYSFQRQNFVBGYVGFEAKDQATPNLVLPYMGFFGSYSQ 819

A related GBS gene <SEQ ID 8767> and protein <SEQ ID 8768> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -8.37

65

570 580 590 600 610 620 630

-1389-

```

2058      2088      2115      2145      2175      2205      2235
ITDKKIPV--DNAADFIKLLMNTAQPIINKQ-SKDGKTPYFVRQOGSGAMNLAKALVTTVATVTGTNDNNADGKLELR
| : : : : | | | | : : : : : | : | | | : : | : | | | :
5 IKKQGLNLSGEELVQFAKNSAMNTSHPVYDTEHTKEIISP---RRQGSGEINVKDAINNTV--EVKAANGNGA---AALK
      650      660      670      680      690      700

2265      2295      2349      2379      2409      2439      2469
ELKEKKFKARILLRNFGKTNKTYIISSEA--IADPVDEKGFRTQNSEHLVSKKADAVTRKVTVEAGKTLAVDLVDYSDA
| : : : : | | | | : : : : : | : : : : | : | | | : | : | :
10 EIG-RQTTFKVTLTNHGKKAQTYAVDNVGGPYTQATEAKSGEITYDTK-IVKGQLTTETPKVTVPQGES--VDVSFTLTLP
      720      730      740      750      760      770      780

2499      2526      2556      2586      2616      2646      2676      2706
EALTRNNFLEGYLNK-DTEGVADLHLPFLGFYGSWTEQKAIDAFEGISEIGNDKRRVQFYVNKETNKTSTFTTNGM
: : | | : | | : : : : : | | : | : | : : : : : | : : : :
15 YSFQRQNFVEGYVGFQAKDQATPNLVLPYMGFFGSYS-QASVSA-PMLYEGGNSNLINTIHSLVGVVMFSNNNDMLGHTGY
      800      810      820      830      840      850

2724      2754      2781      2811      2841      2871      2901      2931
----LSLPIYNNTVFFSPNSP-FYDKAGVRIALRNMEYVQYSIIDPDNKEVRLGRSHDVRKLYRLDYRNSFAMMPDS
| : : | | | : | : | : : | : | : | : | : | : | :
20 EGDDYSKYTDPDLIAISPNGDGSRDYAYPVLFFDRNYKEYTETTTDAQGNK-VKSLGVGKEGTDYSSSSGEWTHSLD
      870      880      890      900      910      920      930

2961      2991      3021      3051      3081      3111
IWDGKIKD*IAGKDKQYIYQIKVQLNNKGVGGDGVQIYQYIYKMDNNKPYLSPDKKTVEKLEDRWK-----
| | | | | | | | | | | | : : : : : | : | : | : | : | : |
25 KWDGTDADGQVVKDQYIY--KVEFT-PAIGGQE-QELNIPVKVDSQAPEVSDLQVTKDGKLRKAKDSSGSLDMTMFVA
      950      960      970      980      990      1000      1010

3159      3189      3219      3249
-----KITFKVQDTGIGLKDVYLQSVKYVGGGNNNLDLITPPGFKK
| : | | |
30 AVNGEEQ---VDGKSWTKLDKDTVQVAENGKVEFKYQDVYGNESKVTTYEVKNIVKEVAAQPELKLTPDGEGKVKAELA
      1520      1530      1540      1550      1560      1570      1580

```

SEQ ID 8768 (GBS362N) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 149 (lane 10; MW 63.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 182 (lane 9; MW 38kDa) and in Figure 149 (lane 11 & 12; MW 38kDa). Purified GBS362N is shown in Figure 235, lanes 3 & 4

GBS362C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 149 (lane 14-16; MW 91kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 18; MW 66.3kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1241

A DNA sequence (GBSx1318) was identified in *S.galactiae* <SEQ ID 3841> which encodes the amino acid sequence <SEQ ID 3842>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.04 Transmembrane 21 - 37 (17 - 38)

----- Final Results -----

bacterial membrane --- Certainty=0.2614(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1390-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA95000 GB:AB042239 PAA [Streptococcus criceti]
Identities = 55/166 (33%), Positives = 81/166 (48%), Gaps = 24/166 (14%)

5   Query: 5   KKTDKFGFRKSKVCRSLCGALLGTVAVVS LATASTEIHADEATTSP TTVTKVPQPVQADT 64
      K+ + FGFRKSK+ +SLCGALLGT VVS+ A A++ TTS T+ DT
      Sbjct: 2   KRKETFGFRKSKISKSLCGALLGTAIVVSV--AGQRALAEDMTTSTTSA-----VDT 51

10  Query: 65   TALNTSKTHSTQATTTTPVEAKENKVVKSETVQSES RV--MPRD-KVVERPETVKASVNS- 120
      TA+ ++T + +A + ++ Q+E + MP D E E VK++ +
      Sbjct: 52   TAVVGTETGNPATNLPEKQADSSSQAEASQAQAEQKTGSM PVDVATTELDEAVKSAEAG 111

      Query: 121 -DVSQPITTTTPPTI-----NEKTVEIPNLAQDTKKVAPKVTVTPE 159
      VSQ T T+ +EK+ EI D K A + +T E
15  Sbjct: 112   VTVSQDETVDKGTGTVGTSQEAEKSGEI---KADYSKQAETIKITTE 154
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3842 (GBS222) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 6; MW 22kDa).

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1242

A DNA sequence (GBSx1319) was identified in *S.agalactiae* <SEQ ID 3843> which encodes the amino acid sequence <SEQ ID 3844>. This protein is predicted to be CylK. Analysis of this protein sequence

25 reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
30      bacterial cytoplasm --- Certainty=0.3738(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1243

A DNA sequence (GBSx1320) was identified in *S.agalactiae* <SEQ ID 3845> which encodes the amino acid sequence <SEQ ID 3846>. This protein is predicted to be CylJ. Analysis of this protein sequence

40 reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45      bacterial cytoplasm --- Certainty=0.1143(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9689> which encodes amino acid sequence <SEQ ID 9690>

50 was also identified.

-1391-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1244

- 5 A DNA sequence (GBSx1321) was identified in *S.agalactiae* <SEQ ID 3847> which encodes the amino acid sequence <SEQ ID 3848>. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0913(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1245

- 20 A DNA sequence (GBSx1322) was identified in *S.agalactiae* <SEQ ID 3849> which encodes the amino acid sequence <SEQ ID 3850>. This protein is predicted to be CylI (fabF). Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
25 INTEGRAL Likelihood = -2.39 Transmembrane 721 - 737 (721 - 738)
INTEGRAL Likelihood = -1.97 Transmembrane 326 - 342 (326 - 343)
INTEGRAL Likelihood = -0.43 Transmembrane 534 - 550 (534 - 550)

30 ----- Final Results -----
bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 9687> which encodes amino acid sequence <SEQ ID 9688> was also identified.

There is also homology to SEQ ID 3852.

A related GBS gene <SEQ ID 8769> and protein <SEQ ID 8770> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: 1.08
GvH: Signal Score (-7.5): -5.97
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
45 ALOM program count: 3 value: -2.39 threshold: 0.0
INTEGRAL Likelihood = -2.39 Transmembrane 712 - 728 (712 - 729)
INTEGRAL Likelihood = -1.97 Transmembrane 317 - 333 (317 - 334)
PERIPHERAL Likelihood = 3.45 492
modified ALOM score: 0.98

-1392-

*** Reasoning Step: 3

----- Final Results -----

5 bacterial membrane --- Certainty=0.1956 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 8770 (GBS361) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 4; MW 84kDa).

10 GBS361-His was purified as shown in Figure 213, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1246

15 A DNA sequence (GBSx1323) was identified in *S.agalactiae* <SEQ ID 3853> which encodes the amino acid sequence <SEQ ID 3854>. This protein is predicted to be CylF. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3766 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1247

30 A DNA sequence (GBSx1324) was identified in *S.agalactiae* <SEQ ID 3855> which encodes the amino acid sequence <SEQ ID 3856>. This protein is predicted to be CylE. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3498 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1248

45 A DNA sequence (GBSx1325) was identified in *S.agalactiae* <SEQ ID 3857> which encodes the amino acid sequence <SEQ ID 3858>. This protein is predicted to be ABC transporter homolog CylB. Analysis of this protein sequence reveals the following:

-1393-

Possible site: 56

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -13.90 Transmembrane 271 - 287 (263 - 291)
 INTEGRAL Likelihood = -10.30 Transmembrane 17 - 33 (14 - 43)
 INTEGRAL Likelihood = -8.60 Transmembrane 114 - 130 (106 - 138)
 INTEGRAL Likelihood = -6.69 Transmembrane 152 - 168 (149 - 178)
 INTEGRAL Likelihood = -1.97 Transmembrane 186 - 202 (185 - 202)

----- Final Results -----

bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9685> which encodes amino acid sequence <SEQ ID 9686> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1249

A DNA sequence (GBSx1326) was identified in *S.agalactiae* <SEQ ID 3859> which encodes the amino acid sequence <SEQ ID 3860>. This protein is predicted to be ABC transporter homolog CylA. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4122(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9683> which encodes amino acid sequence <SEQ ID 9684> was also identified. A further related GBS gene <SEQ ID 8771> and protein <SEQ ID 8772> were also identified. Analysis of this protein sequence reveals homology to membrane protein ABC transporters.

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9085> which encodes the amino acid sequence <SEQ ID 9086>. An alignment of the GAS and GBS sequences follows:

Score = 85.4 bits (208), Expect = 1e-18

Identities = 68/271 (25%), Positives = 129/271 (47%), Gaps = 17/271 (6%)

Query: 39 KGFTEQHVLKDINFDVYKGDFFGIVGRNGSGKSTLLKIISQIYVPEKGQVT--VDGKMVS 96
 K + L+DIN +G F+G++G NG+GK+TL ++ Q + G + VD GK +S
 Sbjct: 10 KKYGSFEALRDINLIFEKGKFGYGLLGPNGAGKTTLFNLLIQNFQKQTSQDQKWEVDGKPLS 69

Query: 97 ----FIELGVGF-----NPELTGRENVYMGAMLGFTKDEVDMDYNDIVDFAEHLHFMNQ 147
 + +G+ F + LT EN+ GA+ G +K +V + D+ + ++ Q
 Sbjct: 70 IKDFYRHIGIVFQSNRLDDNLTVEENLISRGALYGLSKSQVRNRLKDLQTYLDITAIKKQ 129

Query: 148 KLKNYSSGMQVRLAFSVAIKAQGDVLILDEVLA VGDEAFQRCNDYFME-RKDSGKTTIL 206
 K + S G + ++ + A+ Q +L+LDE D +R D + + S T +L
 Sbjct: 130 KYGSLSGGQKRKVDIARALLPQPSLLLLDEPTTGLDPQSRRLWDALQNLNQQSQMTFVVL 189

Query: 207 VTHDMGAVKKYCNRAVLIEDGLVKAYGEPFDVANQYSVDNTETA-EDAMNAEKISVSDIA 265
 +TH + + C+ ++ +G + G+ Q+S N + + +++S++D
 Sbjct: 190 ITHYLEEMSA-CDVLNVLIENIYYSGDIKSFTEQHSTTNLNVVLKPEKSLDQLSIADFV 248

Query: 266 KDLKVSLISNPRITPNDDTITFEVSYEVLKDD 296

-1394-

K ++S I D I+ E +V+ D+
 Sbjct: 249 N--KCQVLSEREIVFKD-ISVBEEMMQVISDN 276

There is also homology to SEQ IDs 358, 482, 644, 686, 1832, 2529, 2720, 3882, 4028, 4104, 4280, 5090,
 5 5498, 6034, 6500.

SEQ ID 8772 (GBS83) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 2; MW 37.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 5; MW 62.6kDa) and in Figure 28 (lane 3; MW 62.6kDa).

10 GBS83-GST was purified as shown in Figure 195, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1250

15 A DNA sequence (GBSx1327) was identified in *S.agalactiae* <SEQ ID 3861> which encodes the amino acid sequence <SEQ ID 3862>. This protein is predicted to be acyl carrier protein homolog AcpC. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3451(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1251

30 A DNA sequence (GBSx1328) was identified in *S.agalactiae* <SEQ ID 3863> which encodes the amino acid sequence <SEQ ID 3864>. This protein is predicted to be CylG (fabG). Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 There is also homology to SEQ ID 3866.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1252

A DNA sequence (GBSx1329) was identified in *S.agalactiae* <SEQ ID 3867> which encodes the amino acid sequence <SEQ ID 3868>. This protein is predicted to be CylD. Analysis of this protein sequence reveals the following:

```

5   Possible site: 60
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2030(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1253

A DNA sequence (GBSx1330) was identified in *S.agalactiae* <SEQ ID 3869> which encodes the amino acid sequence <SEQ ID 3870>. Analysis of this protein sequence reveals the following:

```

20  Possible site: 14
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3219(Affirmative) < succ>
25  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1254

A DNA sequence (GBSx1331) was identified in *S.agalactiae* <SEQ ID 3871> which encodes the amino acid sequence <SEQ ID 3872>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 56
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -8.97    Transmembrane  231 - 247 ( 226 - 251)
      INTEGRAL    Likelihood = -7.06    Transmembrane  141 - 157 ( 134 - 164)
      INTEGRAL    Likelihood = -2.76    Transmembrane   28 - 44 ( 26 - 44)
40  INTEGRAL    Likelihood = -1.38    Transmembrane  123 - 139 ( 121 - 139)
      INTEGRAL    Likelihood = -0.32    Transmembrane  199 - 215 ( 199 - 215)

    ----- Final Results -----
      bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB88836 GB:AL353832 putative integral membrane transport
      protein. [Streptomyces coelicolor A3(2)]
50  Identities = 68/264 (25%), Positives = 123/264 (45%), Gaps = 10/264 (3%)

```

-1396-

Query: 6 RMHFIFIKQYMKQIMEYKIDFFVGVGLVFLTQGLNLLFLNVLFQHIPSLLEGWTFQQIAFI 65
 R + + + M Y+ F + G F L+ + + ++F + +L G++ ++AF+
 Sbjct: 34 RAYGLIAGMWIRSTMAYRTSFALTAFGNFAMTALDFVAILLMFSRVDALGGYSLPEVAFL 93

5
 Query: 66 YGFSLLPKGIDHLFFDNLWALGQRLIRKGEFDKYLTRPISPLFHVIVETQVDALGELLV 125
 YG S + G+ L ++ LG+R +R G D L RP L V + F + LG ++
 Sbjct: 94 YGLSGVSFGLADLAIGSMERLGRR-VRDGTLDTLVLRPAPVLAQVAADRFAALRRGRVVQ 152

10
 Query: 126 GFILL--STTVSSISWTVPKVLLFIFIIPFATLIYTSLSKIATSSIAFWTKQSGAVIYIF- 182
 G ++L + V I WT KVL L + I+ ++ +A + F + + V F
 Sbjct: 153 GLVLGVYALVVVDIDWTAAKVLLLPVALISGAGIFCAVFVAAGAFQFAAQDASEVANAF 212

15
 Query: 183 YMFNDFAKYPVAIYNNLLRWIISFVIPFAFTAYYPAYFLQDRNVYFNIGGVI-----LI 237
 Y +YP ++ L +FV+P AF + PA+Y L R ++ G + L
 Sbjct: 213 YGGTTLQYPPTVFALDLVRGATFVLPLAFVNWLPASYVL-GRPYPLDLPGWVAFTPPLA 271

20
 Query: 238 SLISFMVSLILWHKGVEVYESAGS 261
 + ++ + W G+ Y S GS
 Sbjct: 272 AAACALAGLAWRAGLRSYRSTGS 295

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3873> which encodes the amino acid sequence <SEQ ID 3874>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.86	Transmembrane	227 - 243 (225 - 251)
INTEGRAL	Likelihood = -7.22	Transmembrane	141 - 157 (133 - 164)
INTEGRAL	Likelihood = -6.37	Transmembrane	123 - 139 (114 - 140)
INTEGRAL	Likelihood = -2.97	Transmembrane	26 - 42 (26 - 49)

30
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35
 The protein has homology with the following sequences in the databases:

>GP:CAB88836 GB:AL353832 putative integral membrane transport protein. [Streptomyces coelicolor A3(2)]
 Identities = 69/262 (26%), Positives = 125/262 (47%), Gaps = 10/262 (3%)

40
 Query: 8 HAIFIKQYLKQIMEYKIDFFVGVGLVFLTQGLNLLFLSVLFQHIPSLLEGWTFEQIAFIY 67
 + + + M Y+ F + G F L+ + + ++F + +L G++ ++AF+YG
 Sbjct: 36 YGLIAGMWIRSTMAYRTSFALTAFGNFAMTALDFVAILLMFSRVDALGGYSLPEVAFLY 95

45
 Query: 68 FCLIPKGIDHLFFDNLWALGQRLVRKGEFDKYLTRPISPLFHVIVETQVDALGELLVGV 127
 + G+ L ++ LG+R VR G D L RP L V + F + LG ++ G+
 Sbjct: 96 LSGVSFGLADLAIGSMERLGRR-VRDGTLDTLVLRPAPVLAQVAADRFAALRRGRVVQGL 154

50
 Query: 128 ILL--VTTAGSIVWTLPKVLLFILVIPFATLIYTSLSKIATASISFWTKQSGAVIYIF-YM 184
 ++L I WT KVL L + + I+ ++ +A + F + + V F Y
 Sbjct: 155 LVLGYALVVVDIDWTAAKVLLLPVALISGAGIFCAVFVAAGAFQFAAQDASEVANAFY 214

55
 Query: 185 FNDFSKYPMSIYHSFLRWLISFIIPFAFTAYYPASYFLTQHLLFNIGGLV-----VVSL 239
 +YP +++ L +F++P AF + PASY L G+ ++ G V + +
 Sbjct: 215 GTTTLQYPPTVFALDLVRGATFVLPLAFVNWLPASYVL-GRPYPLDLPGWVAFTPPLAAA 273

60
 Query: 240 LVLALSLLKLWKWGLDAYESAGS 261
 AL+ W+ GL +Y S GS
 Sbjct: 274 ACCALAGLAWRAGLRSYRSTGS 295

An alignment of the GAS and GBS proteins is shown below.

Identities = 208/261 (79%), Positives = 238/261 (90%)

Query: 1 MTKYQRMHFIFIKQYMKQIMEYKIDFFVGVGLVFLTQGLNLLFLNVLFQHIPSLLEGWTFQ 60

-1397-

M K + MH IFIKQY+KQIMEYK+DF VGVLG VFLTQGLNLLFL+VLFQHIPSLGWTTF+
 Sbjct: 1 MAKLRCHMAIFIKQYLKQIMEYKVDFVVGVLGVFLTQGLNLLFLSVLFLQHIPSLGWTTFE 60
 Query: 61 QIAFIYGFSLLPKGIDHLFFDNLWALGQRLIRKGEFDKYLTRPISPLFHVIVETTFQVDAL 120
 5 QIAFIYGF L+PKGIDHLFFDNLWALGQRL+RKGEFDKYLTRPISPLFHVIVETTFQVDAL
 Sbjct: 61 QIAFIYGFCLIPKGIDHLFFDNLWALGQRLVRKGEFDKYLTRPISPLFHVIVETTFQVDAL 120
 Query: 121 GELLVGFILLSTTVSSISWTVPKVLLFIFIFIIPFATLIYTSLKATSSIAFWTKQSGAVIY 180
 GELLVG ILL TT SI WT+PKVLLFI +IPFATLIYTSLKAT+SI+FWTKQSGAVIY
 10 Sbjct: 121 GELLVGVILLVTTAGSIVWTLPKVLLFILVIPFATLIYTSLKATASISFWTKQSGAVIY 180
 Query: 181 IFYMFNDFAKYPVAIYNLLRWIISFVIPFAFTAYYPAAYFLQDRNVYFNIGGVILISLI 240
 IFYMFNDF+KYP++IY++ LRW+ISF+IPFAFTAYYPA+YFL +++ FNIGG++++SL+
 15 Sbjct: 181 IFYMFNDFS KYPMSIYHSFLRWLISFIPFAFTAYYPASYFLTGQHLLFNIGGLVVVSL 240
 Query: 241 SFMVSLILWHKGVEVYESAGS 261
 +SL LW G++ YESAGS
 Sbjct: 241 VLALSLKLWKWGLDAYESAGS 261

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1255

A DNA sequence (GBSx1332) was identified in *S.agalactiae* <SEQ ID 3875> which encodes the amino acid sequence <SEQ ID 3876>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -15.60 Transmembrane 147 - 163 (134 - 178)
 INTEGRAL Likelihood = -8.55 Transmembrane 119 - 135 (114 - 141)
 INTEGRAL Likelihood = -7.86 Transmembrane 238 - 254 (235 - 260)
 30 INTEGRAL Likelihood = -1.70 Transmembrane 215 - 231 (212 - 231)
 INTEGRAL Likelihood = -1.06 Transmembrane 61 - 77 (61 - 77)
 INTEGRAL Likelihood = -0.22 Transmembrane 27 - 43 (27 - 43)
 ----- Final Results -----
 35 bacterial membrane --- Certainty=0.7241(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB88837 GB:AL353832 putative integral membrane protein.
 [Streptomyces coelicolor A3(2)]
 Identities = 60/271 (22%), Positives = 118/271 (43%), Gaps = 13/271 (4%)
 Query: 6 RRYKPFISTGIQGLITYRVDFILYRIGDVIGAFVAFYLWKAVFDSSSQSLIQGFQLSDMI 65
 R Y + G + TYR + + + Y + A++D Q + G+ + +
 45 Sbjct: 7 RLYVAVAAGGFRRYATYRAATAAGVFTNTVFGILLVYTYLALWDEKPQ--LGGYDQAQAV 64
 Query: 66 LYIIMS-FVTNLLTRTDSSFM--IGDEVKDGSIIMRLRPVHFAASYLFMEIGSRWLIFL 122
 ++ + + L F + + ++ G + + L RP +L ++G L
 50 Sbjct: 65 TFWLWGQALLAALAIGGGGFEDLMERIRTGDVAVDLYRPADLQLWWLAADVGRAVFQLL 124
 Query: 123 SIGV-PFLLVITGVRLFLGTDLIQAIVLVVFIYIISIILAFLINFFFNICFGFSAFVFKNL 181
 GV PF+ LF L + + + +++++LA ++ F SAF +
 55 Sbjct: 125 GRGVVPFVFG----SLFFPVLPREVSVWAAFLVAVVLAMVVGFAALRYLVALSAFWLLDG 180
 Query: 182 WGSNLLKNSLVAFMSGSLIPLTFFPKIVADILGFLPFSSLIYTPVMIIGKYDGSQIVQA 241
 G + F SG L+PL FP ++ D++ LP+SSL+ P +++G+ D +
 Sbjct: 181 TGVTDMAWLAGLFCSGMLLPLNVFPGVLGDVVRALPWSSLLQGPADVLLGEADP---LGT 237
 60 Query: 242 LLLQIFWLIWMVALSQLIWKKVQLHITIQGG 272
 L Q W + ++AL +L+ + +QGG
 Sbjct: 238 YLFQASAWALLAIGRLVQSAATTRRVVQGG 268

-1398-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3877> which encodes the amino acid sequence <SEQ ID 3878>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -9.18    Transmembrane  252 - 268 ( 248 - 277)
INTEGRAL    Likelihood = -7.22    Transmembrane  161 - 177 ( 151 - 187)
INTEGRAL    Likelihood = -6.10    Transmembrane  133 - 149 ( 128 - 160)
INTEGRAL    Likelihood = -2.81    Transmembrane  213 - 229 ( 211 - 230)

----- Final Results -----
      bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF11144 GB:AE002002 conserved hypothetical protein [Deinococcus radiodurans]
Identities = 56/268 (20%), Positives = 113/268 (41%), Gaps = 21/268 (7%)

```

```

Query: 15  MWSFWKRYRPFLSAGIQELITYRVNFFLYRIGDVMGAFVAYYLWKAVFDSSKQSLINGFT 74
           M +FW++ R  +  +  + YR  ++ +  +  V  +W  S+  ING+T
Sbjct: 1   MTNFWRKVRVLWAVSLASTLEYRAETIIWMLSGTLN-LVMMLVWMTQAKSAPGGQINGYT 59

Query: 75  LSDMTFYIIMSFTVTTLLTKSDSSFMIGEEVKDGSIIIMRLLRPV-----HFAASYLFMEIG 129
           Y + +++ + L      + +  +++ G++  LL P+  FAA  +
Sbjct: 60  PQAFAGYFLATWLVSQQLLVVWVGWELDYKIRQGTLSPELLHPIDPLWREFAAH--LTDKA 117

Query: 130 FRWIVLMSVGFPPFLMVLSGIKVMAGLSILQVLASSCLYLVSLLLAFL---INFYFNICFG 186
           FR      P ++VL  ++ A L+  Q  +  Y  L LA L  + F +  G
Sbjct: 118 FR-----LPIMLVL--LLIFAALTGAQFTSQWWAYPAVLGLALLGLCVRFLWEYTLG 167

Query: 187 SSAFVFKNLWGSNLLKNALVAFMSGSLIPLAFFPKMVSIVLSFLPFSSLVYTPVMIVIGK 246
           AF ++      +  A  G  PL+F+P  +  + ++ PF ++  P ++ GK
Sbjct: 168 LLAFTWESSSSFGFVWLFYAAFGGMFAPLSFYPGWLQTLAAWTPFFYMLGLPAALLAGK 227

Query: 247 YSLSQIMVALSIQIFWLLVMVLSQVIW 274
           S ++ +  + + WL VM ++ + +W
Sbjct: 228 ASGAEALRGAGVLLGLWLAVMMLVRRWVW 255

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 199/268 (74%), Positives = 236/268 (87%)

Query: 5   WRRYKPFISTGIQGLITYRVDFILYRIGDVIGAFVAFYLWKAVFDSSSQSLIQGFQLSDM 64
           W+RY+PF+S GIQ LITYRV+F LYRIGDV+GAFVA+YLWKAVFDSS QSLI GF LSDM
Sbjct: 19  WKRYRPFLSAGIQELITYRVNFFLYRIGDVMGAFVAYYLWKAVFDSSKQSLINGFTLSDM 78

Query: 65  ILYIIMSFTVNTLLTRTDSSFMIGDEVKDGSIIMRLLRPVHFAASYLFMEIGSRWLIFLSI 124
           YIIMSFTV LLT++DSSFMIG+EVKDGSIIMRLLRPVHFAASYLFMEIG RW++ +S+
Sbjct: 79  TFYIIMSFTVTTLLTKSDSSFMIGEEVKDGSIIIMRLLRPVHFAASYLFMEIGFRWIVLMSV 138

Query: 125 GVPFLLVITGVRLFLGTDLIQAIVLVVFYIISIILAFLINFFFNICFGFSAFVFKNLWGS 184
           G PFL+V++G+++ G  ++Q +  Y++S++LAFLINF+FNICFG SAFVFKNLWGS
Sbjct: 139 GPFPLMVLSGIKVMAGLSILQVLASSCLYLVSLLLAFLINFFFNICFGSSAFVFKNLWGS 198

Query: 185 NLLKNLSLVAFMSGSLIPLTFFPKIVADILGFLPFSSLIYTPVMIIGKYDGSQIVQALLL 244
           NLLKN+LVAFMSGSLIPL FFPK+V+ +L FLPFSSL+YTPVMI+IGKY SQI+ AL L
Sbjct: 199 NLLKNALVAFMSGSLIPLAFFPKMVSIVLSFLPFSSLVYTPVMIVIGKYSLSQIMVALSL 258

Query: 245 QIFWLIVMVALSQLIWKKVQLHITIQQG 272
           QIFWL+VMV LSQ+IWKKVQ H+TIQQG
Sbjct: 259 QIFWLLVMVLSQVIWKKVQYHLITIQQG 286

```

-1399-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1256

A DNA sequence (GBSx1333) was identified in *S.agalactiae* <SEQ ID 3879> which encodes the amino acid sequence <SEQ ID 3880>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2013(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9681> which encodes amino acid sequence <SEQ ID 9682> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF09790 GB:AE001882 ABC transporter, ATP-binding protein
[Deinococcus radiodurans]

Identities = 141/331 (42%), Positives = 201/331 (60%), Gaps = 34/331 (10%)

Query: 10 MIEVSHLQKNFIKTVKAPGLKGAQSFRLRPEKHTFEAVKDLTFDVPKGQILGFIGANGAG 69
MIEV HL K+F + AV+D++F +P G+I+G++G NGAG
Sbjct: 46 MIEVRHLCKSFARK-----PAVQDISFSIPAGEIVGYLGPNGAG 84

Query: 70 KSTTIKMLTGILKPTSGFCRIDGKLPQENRQNYVKDIGVVFQRTQLWWDLALQETYTVL 129
KSTTIK+LTG+L P SG R+ G +P + R+ +V +G VFGQRT LWWDL ++E+ +L
Sbjct: 85 KSTTIKVLTLGLVPDSGEVRVGGGLVPWKQRRQHVRLGAVFGQRTTLWWDLPVRESLELL 144

Query: 130 KEIYDVPDKFEFRKMAFLNEVLELNDFIKDPVRTLSLGQMRADIAASLLHNPVLFLE 189
+ +Y VP F + +A E+LEL F+ P R LSLGQMRAD+AA+LLH+P++LFLDE
Sbjct: 145 RHVYRVPAARFAENLAGFTELELGPFLNTPARALSLGQMRADLAAALLHDPPELLFLDE 204

Query: 190 PTIGLDVSVKDNIRRAITQINQEEETTILLTTHDLSDIEQLCHRIFMIDRGQEIFDGTVS 249
PT+GLDV K+ IR + +N E T+LLTTHDL D+E+L R+ MID G+ +FDG ++
Sbjct: 205 PTVGLDVVAKERIREFVKAVNAERGVTVLLTTHDLGDVERLARRVMMIDTGRLLFDGPLA 264

Query: 250 QLKETFGKMKTL--SFDLRPGQEHIS--SLIGKSEINIKRNDLVLDIQYDSSRYQTADII 306
+L+ +G + L F+ P Q + +L+G+ ++ Y S A I
Sbjct: 265 ELQARYGGERELWVEFEKAPAPALPGLTLLGQDGPVR-----YGFSGAAAAPIA 315

Query: 307 QQTLDVSVRDLKMTDADIEDIIRRFYRNL 337
Q T A VRDL + + ++E IRR Y L
Sbjct: 316 QVT-ALAPVRDLAVKEPEVEATIRRIYEGNL 345

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3881> which encodes the amino acid sequence <SEQ ID 3882>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3315(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 272/330 (82%), Positives = 305/330 (92%)

-1400-

Query: 8 MSMIEVSHLQKNFIKTVKAPGLKGAFQSFLRPEKHTFEAVKDLTFDVPKGQILGFIGANG 67
 M MIEVSHLQKNF KT+K PGLKGA +SF+ P + FEAVKDL+F+VPKGQILGFIGANG
 Sbjct: 1 MVMIEVSHLQKNFSKTIKEPGLKGALSKFSVHPPREIFEAVKDLSFEVPGQILGFIGANG 60

Query: 68 AGKSTTIKMLTGILKPTSGFCRIDGKLPQENRQNYVKDIGVVFQRTQLWWDLALQETTYT 127
 AGKSTTIKMLTGILKPTSG+CRI+GK+PQ+NRQ YV+DIG VFGQRTQLWWDLALQETTY
 Sbjct: 61 AGKSTTIKMLTGILKPTSGYCRINGKIPQDNRYVVRDIGAVFGQRTQLWWDLALQETTYV 120

Query: 128 VLKEIYDVPDKEFRKRMAFLNEVLELNDFIKDPVRTLSLGQMRADIAASLLHNPVKVLF 187
 VLKEIYDVP+K FRKRM FLNEVL+LN+FIKDPVRTLSLGQMRADIAASLLHNPVKVLF
 Sbjct: 121 VLKEIYDVPEKAFRKRMDFLNEVLDLNEFIKDPVRTLSLGQMRADIAASLLHNPVKVLF 180

Query: 188 DEPTIGLDVSVKDNIRRAITQINQEEETTILLTTHDLSDIEQLCHRIFMIDRGQEIFDGT 247
 DEPTIGLDVSVKDNIRRAITQINQEEETTILLTTHDLSDIEQLC RI MID+GQEIFDGT
 Sbjct: 181 DEPTIGLDVSVKDNIRRAITQINQEEETTILLTTHDLSDIEQLCDRIIMIDKGQEIFDGT 240

Query: 248 VSQKETFQKMKTL SFDLRPGQEHSSSLIGKSEINIKRNDLVLDIQYDSSRYQTADIIQ 307
 V+QLK++FGKMK+LSF+L+PGQE + S +G +I ++R++L LDIQYDSSRYQTADIIQ
 Sbjct: 241 VTQLKQSFQKMKSLSFELKPGQEQVVSQFMGLPDITVERHELSDIQYDSSRYQTADIIQ 300

Query: 308 QTLADFSVRDLKMTDADIEDIIRRFYRNEL 337
 +T+ADF+VRD+KMTD DIEDI+RRFYR EL
 Sbjct: 301 KTMADFAVRDVKMTDVEDIEDIVRRFYRKEL 330

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1257

A DNA sequence (GBSx1334) was identified in *S.agalactiae* <SEQ ID 3883> which encodes the amino acid sequence <SEQ ID 3884>. This protein is predicted to be Fmt. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.39 Transmembrane 21 - 37 (8 - 39)
 INTEGRAL Likelihood = -7.75 Transmembrane 360 - 376 (359 - 381)

----- Final Results -----
 bacterial membrane --- Certainty=0.4758(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8775> which encodes amino acid sequence <SEQ ID 8776> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: 8.85
 GvH: Signal Score (-7.5): -3.75
 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 2 value: -9.39 threshold: 0.0
 INTEGRAL Likelihood = -9.39 Transmembrane 21 - 37 (8 - 39)
 INTEGRAL Likelihood = -7.75 Transmembrane 353 - 369 (352 - 374)
 PERIPHERAL Likelihood = 4.24 92
 modified ALOM score: 2.38

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.4758(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1401-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA24012 GB:AB009635 Fmt [Staphylococcus aureus]
Identities = 72/279 (25%), Positives = 125/279 (43%), Gaps = 25/279 (8%)

5 Query: 49 LHRFMRKNNVNGMMIVSDNTGKPIITISHGINRGEVETDIEN--NKLFPMASLQKLMTGII 106
+ +++++ + NG + + +N GK + +S G + E I+N N +F + S QK TG++
Sbjct: 79 IDKYLQSSLFNGSVAIYEN-GK-LKMSKGYGYQDFEKGIKNTPTMTFLIGSAQKFSTGLL 136

10 Query: 107 IQRLIDQDVLSEDDRSLQFFPQVKGNSITIHQLLTHTSGLREKGVKVSPLYKNEREQLQ 166
+++L ++ ++ +D +S++ P K S I + L+ H SGL + K S KN + ++
Sbjct: 137 LKQLEEEHKININDPVSKYLPWFKTSKPIPLKDLMLHQSGLYK--YKSSKDYKNLDQAVK 194

15 Query: 167 FCLKHYNFVNK-KSWYYSNINFSFLTGIATQVTGRTYAEVLDDVIKNPLRLDDTQSYQSV 225
K K K Y++ N+ L + +VTG++YAE I +PL+L T Y
Sbjct: 195 AIQKRGRIDPKKYKKHMYNDGNLVLAKVIEEVTGKSYAENYYTKIGDPLKLQHTAFYD-- 252

20 Query: 226 VNHDLVSPMRKNGKLNKINIF----NQVSTAYAGDFFTTPLNFWVLMRSFSKGYFFPT- 280
+ K N + N + YGAG+ + TP + L+ + F
Sbjct: 253 -EQPFKKYLAKGYAYNSTGLSFLRPNILDQYYGAGNLYMTPTDMGKLITQIQYKLFSPK 311

25 Query: 281 -----DEYTKHQNDASHYYGGLYMHGRIVNSNGTFF 312
+ TK D Y G Y + NG FF
Sbjct: 312 ITNPLLHEFGTKQYPD---EYRYGFYAKPTLNRLNGGFF 347

25 There is also homology to SEQ ID 3886.

A related GBS gene <SEQ ID 8773> and protein <SEQ ID 8774> were also identified. Analysis of this protein sequence reveals the following:

30 Lipop: Possible site: -1 Crend: 6
McG: Discrim Score: 14.89
GvH: Signal Score (-7.5): -3.75
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 1 value: -9.39 threshold: 0.0
INTEGRAL Likelihood = -9.39 Transmembrane 14 - 30 (1 - 32)
35 PERIPHERAL Likelihood = 4.24 85
modified ALOM score: 2.38

*** Reasoning Step: 3

40 ----- Final Results -----
bacterial membrane --- Certainty=0.4758(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the databases:

29.6/49.6% over 218aa
Bacillus cereus
GP|4127525| D-stereospecific peptide hydrolase Insert characterized

50 ORF00162(478 - 1083 of 1644)
GP|4127525|emb|CAA09676.1||AJ011526(67 - 285 of 389) D-stereospecific peptide hydrolase
{Bacillus cereus}
%Match = 5.8
%Identity = 29.5 %Similarity = 49.5

55 Matches = 62 Mismatches = 96 Conservative Sub.s = 42

330	360	390	420	450	480	510	540
MILRRLFMVRKFLKSLLSLFLIAVIATGISVACFFFI	PENKGNITPILLHRFMRKNNVNGMMIVSDNTGKPI	TISHGINR					
:::	:		: :	:	:		::: : :
60 TCASLALLIAGSSLLYTTPTSIVKAEPTQNVSSSLQ	TNTQDRDTSVKQAMRDTLQLGYPGILAKTSEGGKT	WGYAAGIAD					
20	30	40	50	60	70	80	
570	600	630	660	705	735	753	

-1402-

```

5      GEVETDIENKLFPMASLQKLMGTGIIQRLIDQDVLSEDDRLSQFFPQV---KG--SNSITIHQLLTHTSGL----REKG
      :  ::  :  | :  | :  |  ::  :  :  |  ||  :  ::  |  |  |||:::  |  |||:  |  |
      LRTKKPMKTDFFRIGSVTKTFTATVVLQLVGENRLKLDHIEDWLPQVIGNGYDGNKITIQEILNHTSGIAEYSRSKD
      100      110      120      130      140      150      160

10      807      834      864      894      924      954      978
      VKVSPYLKN--EREQLQFCLKHY-NFVNKKSQWYYSNINFSFLTGIATQVTGRTYAELVDDVIKNPLRLDDT--QSYQSVV
      |  :  | :  |  :  :  :  |  |  |||  :  :  :  :  |||  :  |||  | :  |  |  |  :  |  ||:
      VDFTDTKKSQYTAEEVLVKMGISFPPDFAPGKGWSYSNTGYVLLGILIEKVTGNSYAEVEENRIIEPLELSNTFLPGNSSVI
      180      190      200      210      220      230      240

15      993      1023      1053      1083      1113      1143      1173      1203
      ---NH--DLVSPMRKNGKLNKINIFNQVSTAYGAGDFFTTPLNFWVLMRSFSKGYFFPTDEYTKHONDAISHYYGGLYMH
      ||  |  |  :  :  :  |  |||  :  :  :  :  :  |  :  :  :  :  :  |
      PGTNHARGYVQP-DGASELKDVITYN-PSAGSSAGDMISTADDLNKFFSYLLGGKLLKEQQLKQMLTTVPTGKEGIDGYG
      260      270      280      290      300      310      320

```

SEQ ID 8776 (GBS61) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 3; MW 68kDa).

GBS61-GST was purified as shown in Figure 195, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1258

A DNA sequence (GBSx1335) was identified in *S.agalactiae* <SEQ ID 3887> which encodes the amino acid sequence <SEQ ID 3888>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

```

```

30      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2398(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1259

A DNA sequence (GBSx1336) was identified in *S.agalactiae* <SEQ ID 3889> which encodes the amino acid sequence <SEQ ID 3890>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood = -5.57      Transmembrane      16 - 32 ( 13 - 33)

```

```

45      ----- Final Results -----
      bacterial membrane --- Certainty=0.3230(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1403-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1260

A DNA sequence (GBSx1337) was identified in *S.agalactiae* <SEQ ID 3891> which encodes the amino acid sequence <SEQ ID 3892>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3910(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1261

A DNA sequence (GBSx1338) was identified in *S.agalactiae* <SEQ ID 3893> which encodes the amino acid sequence <SEQ ID 3894>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4239(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1262

A DNA sequence (GBSx1339) was identified in *S.agalactiae* <SEQ ID 3895> which encodes the amino acid sequence <SEQ ID 3896>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4349(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1263

A DNA sequence (GBSx1340) was identified in *S.agalactiae* <SEQ ID 3897> which encodes the amino acid sequence <SEQ ID 3898>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4962 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1264

A DNA sequence (GBSx1341) was identified in *S.agalactiae* <SEQ ID 3899> which encodes the amino acid sequence <SEQ ID 3900>. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4014 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG38044 GB:AF295925 Orf28 [Streptococcus pneumoniae]
Identities = 23/35 (65%), Positives = 28/35 (79%)

Query: 9 LIHWEGNSGDKLIEHQTSATGWYYQVDRSFSQPKG 43
L +WEGNSGDKL+E QT AT WYYQ+++ FSQ G
Sbjct: 180 LTYWEGNSGDKLLERQTRATEWYYQIEKGFSTQNG 214

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1265

A DNA sequence (GBSx1342) was identified in *S.agalactiae* <SEQ ID 3901> which encodes the amino acid sequence <SEQ ID 3902>. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2036 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1266

A DNA sequence (GBSx1343) was identified in *S.agalactiae* <SEQ ID 3903> which encodes the amino acid sequence <SEQ ID 3904>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10933> which encodes amino acid sequence <SEQ ID 10934> was also identified.

SEQ ID 3904 (GBS153) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 25 (lane 3; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 4; MW 47kDa).

GBS153-GST was purified as shown in Figure 198, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1267

A DNA sequence (GBSx1344) was identified in *S.agalactiae* <SEQ ID 3905> which encodes the amino acid sequence <SEQ ID 3906>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2036(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1268

A DNA sequence (GBSx1345) was identified in *S.agalactiae* <SEQ ID 3907> which encodes the amino acid sequence <SEQ ID 3908>. Analysis of this protein sequence reveals the following:

Possible site: 19

-1406-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2570(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA59773 GB:X85787 tasA [Streptococcus pneumoniae]
 Identities = 18/33 (54%), Positives = 28/33 (84%)

Query: 2 DVQSDENFAFKIFKVAKAGLSLDVFDKLVGRF 34
 + QSD+N F++FKV+K KG++LD FD+++GRF
 Sbjct: 320 EYQSDKNPFFVEVFKVSKTKGIALDPFDEIIGRF 352

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3909> which encodes the amino acid
 sequence <SEQ ID 3910>. Analysis of this protein sequence reveals the following:

Possible site: 56

20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2405(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 18/34 (52%), Positives = 25/34 (72%)

30 Query: 1 MDVQSDENFAFKIFKVAKAGLSLDVFDKLVGRF 34
 +DVQSDE+F FK+ KV K+KG+ L+ D+ V F
 Sbjct: 31 LDVQSEDEFGFKVVKVLKSKGIVLNALDESVCGF 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 **Example 1269**

A DNA sequence (GBSx1346) was identified in *S.agalactiae* <SEQ ID 3911> which encodes the amino acid sequence <SEQ ID 3912>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

Possible site: 52

40 >>> Seems to have no N-terminal signal sequence

 INTEGRAL Likelihood = -1.17 Transmembrane 169 - 185 (168 - 185)

----- Final Results -----

45 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAC13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 53/109 (48%), Positives = 75/109 (68%)

55 Query: 13 IPKINQDLPIYAGSEEDNLQRGVGHLEGISLPFGGASTHAVLSGQRGMPAARLFADLDKM 72
 IP I+ DLP+Y G+ +D L +G+GHLEG SLP+GG T +V++G RG+ A +F +LDK+
 Sbjct: 93 IPSISLDLPVYHGTADDTLLKGLGHLEGTSLPVGGEGRSVITGHRGLAEATMFTNLDDKV 152

Query: 73 KKGDYFYVTVNLKETLAYQVDRIMVIEPSQLDAVSIEEDKDYVTLTCTP 121

-1407-

K G D V E L Y+V V+EP + +A+ +EE KD +TL+TCTP
 Sbjct: 153 KTGDSLIVEVFGEVLTYRVTSTKVVEPEETEALRVEEGKDLLTLVTCTP 201

There is also homology to SEQ ID 3740 and to SEQ ID 3910.

- 5 SEQ ID 3912 (GBS194) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 177 (lane 2; MW 24kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1270

- 10 A DNA sequence (GBSx1347) was identified in *S.agalactiae* <SEQ ID 3913> which encodes the amino acid sequence <SEQ ID 3914>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.15 Transmembrane 880 - 896 (876 - 898)
 15 INTEGRAL Likelihood = -4.78 Transmembrane 24 - 40 (23 - 42)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8777> which encodes amino acid sequence <SEQ ID 8778> was also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: -1 Crend: 8
 SRCFLG: 0
 McG: Length of UR: 20
 Peak Value of UR: 2.80
 Net Charge of CR: 5
 McG: Discrim Score: 10.81
 30 GvH: Signal Score (-7.5): -3.76
 Possible site: 29
 >>> Seems to have an uncleavable N-term signal seq
 Amino Acid Composition: calculated from 1
 ALOM program count: 2 value: -5.15 threshold: 0.0
 35 INTEGRAL Likelihood = -5.15 Transmembrane 867 - 883 (863 - 885)
 INTEGRAL Likelihood = -4.78 Transmembrane 11 - 27 (10 - 29)
 PERIPHERAL Likelihood = 7.58 531
 modified ALOM score: 1.53
 icm1 HYPID: 7 CFP: 0.306
 40 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 LPXTG motif: 859-863

- 50 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8778 (GBS104) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 5; MW 95kDa).

GBS104-His was purified as shown in Figure 221, lane 9-10.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1271

A DNA sequence (GBSx1348) was identified in *S.agalactiae* <SEQ ID 3915> which encodes the amino acid sequence <SEQ ID 3916>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -15.28    Transmembrane  257 - 273 ( 252 - 280)
10    INTEGRAL    Likelihood =  -7.11    Transmembrane   19 - 35 ( 16 - 39)

----- Final Results -----
          bacterial membrane --- Certainty=0.7114(Affirmative) < succ>
          bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
15          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC13546.GB:AF019629 putative fimbria-associated protein
  [Actinomyces naeslundii]
20  Identities = 79/178 (44%), Positives = 112/178 (62%), Gaps = 7/178 (3%)

Query: 65  RIALANAYNETLSRNPLL-----IDPFTSKQKEGLREYARMLLEVHEQ--IGHVAIPSIGV 117
      ++  A+AYN+ LS  +L      +      K+  +YA +L+ + + + + IPSI +
Sbjct: 39  QVEQAHAYNDALSAGAVLEANNHVPTGAGSSKSSLOQYANILKANNEGLMARLKIPSISL 98

Query: 118 DIPIYAGTSETVLQKGSGLHLEGTSLPVGGLSTHSLTAHRLPTARLFTDLNKVKKGQIF 177
      D+P+Y GT++  L KG GHLEGTSLPVG  T SV+T HRGL  A +FT+L+KVK G
Sbjct: 99  DLPVYHGTADDTLLKGLGHLEGTSLPVGGEGRSVITGHRGLAEATMFTNLDKVKTGDSL 158

Query: 178 YVTNIKETLAYKVVSIKVVDPALTSEVKIVNGKDYITLLTCTPYMINSHRLLVKGERI 235
      V   E L Y+V S KVV+P      +++ GKD +TL+TCTP  IN+HR+L+ GERI
30  Sbjct: 159 IVEVFGEVLTYRVSTSKVVEPEETEALRVERGKDLLTLVTCTPLGINTHRILLTGERI 216

```

There is also homology to SEQ ID 3740.

SEQ ID 3916 (GBS208) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 5; MW 35kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 8; MW 59.7kDa) and in Figure 160 (lane 5; MW 60kDa).

GBS208-GST was purified as shown in Figure 224, lane 7-8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1272

A DNA sequence (GBSx1349) was identified in *S.agalactiae* <SEQ ID 3917> which encodes the amino acid sequence <SEQ ID 3918>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood =  -9.13    Transmembrane  265 - 281 ( 260 - 284)
50  ----- Final Results -----

```

-1409-

bacterial membrane --- Certainty=0.4652 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 96/265 (36%), Positives = 150/265 (56%), Gaps = 10/265 (3%)

10 Query: 41 QASHANINAFKEAVTKIDRVEINRRIELAYAYNASI-AGAKTNGEYPALKDPYSAEQKQA 99
 Q + + + A A R + ++E A+AYN ++ AGA P A +
 Sbjct: 15 QYNQSKVTADYSAQVDGARPDAKTQVEQAHAYNDALSAGAVLEANNHV---PTGAGSSKD 71

15 Query: 100 GVVEYARMLEVKQ--IGHVLIIPRINQDIPIYAGSAEENLQRGVGHLEGTSLPVGGESTH 157
 ++YA +L+ + + + IP I+ D+P+Y G+A++ L +G+GHLEGTSLPVGGE T
 Sbjct: 72 SSLQYANILKANNEGLMARLKIPISLDELPHYHTADDTLKGLGHLEGTSLPVGGEETR 131

20 Query: 158 AVLTARHGLPTAKLFTNLDKVTVGDRFYIEHIGGKIAYQVDQIKVIAPDQLEDLYVIQGE 217
 +V+T HRGL A +FTNLDKV GD +E G + Y+V KV+ P++ E L V +G+
 Sbjct: 132 SVITGHRGLAEATMFTNLDKVKTDGSLIVEVFGEVLTyrvtstkvvepeetealrveegk 191

25 Query: 218 DHVTLLTCTPFYMINSHRLLVRGKRI-PYVEKTVQKDSKTFRQQYLYTAMVWVVLILLS 276
 D +TL+TCTP IN+HR+L+ G+RI P K + K + +A+ + GLI++
 Sbjct: 192 DLLTLVTCTPLGINTHRILLTGERIYPTPAKDLAAGKRDPVPHFPWWAVGLAAGLIVVG 251

Query: 277 LLIW---FKKTKQKKRRKNEKAASQ 298
 L +W + + K+R A+Q
 Sbjct: 252 LYLWRSGYAAARAKERALARARAAQ 276

30 There is also homology to SEQ ID 3740.

SEQ ID 3918 (GBS209) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 4; MW 62kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 3; MW 37.2kDa).

GBS209-His was purified as shown in Figure 221, lane 8.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1273

A DNA sequence (GBSx1350) was identified in *S.galactiae* <SEQ ID 3919> which encodes the amino acid sequence <SEQ ID 3920>. Analysis of this protein sequence reveals the following:

40 Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -9.66 Transmembrane 281 - 297 (276 - 300)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:BAB04080 GB:AP001508 unknown [Bacillus halodurans]
 Identities = 45/141 (31%), Positives = 63/141 (43%), Gaps = 20/141 (14%)

55 Query: 153 TGELDLLKVGVDGDTKKPLAGVVFELYEKNRTPIRVKNGVHSQDIDAAKHLETDSSGHI 212
 TG L++ KV D DT + L G F LY+ G IR LET G
 Sbjct: 1084 TGSLEVTKV--DADTGEVLQGATFTLYDSEGEFAIRT-----LETGEDGKA 1127

-1410-

Query: 213 RISGLIHGDYVLKEIETQSGYQIGQAETAVTIEKSKTVTVTIENKKVPTPKVPSRGGL-I 271
 L++GDY+LKE GY +G +T + VT+EN+K +V + G + +
 Sbjct: 1128 TFFNLLLYGDYLLKEDSAPEGYLVGINDTQRVTIDTVLHEVTVENEKSDINRVSAVGAVQL 1187

Query: 272 PKTGEQQAMALVIIGGILIAL 292
 K E+ +L G L AL
 Sbjct: 1188 QKVDEETGESL---QGALFAL 1205
 Identities = 64/259 (24%), Positives = 113/259 (42%), Gaps = 48/259 (18%)

Query: 16 GTMFGISQT---VLAQETHQLTIVHLEARDIDRPNP----QLEIAPKE-GTPIEGVLYQL 67
 G + GI+ T + H++T+ + E DI+R + QL+ +E G ++G L+ L
 Sbjct: 1147 GYLVGINDTQRVTIDTVLHEVTVEN-EKSDINRVSAVGAVQLQKVDEETGESLQGALFAL 1205

Query: 68 YQLKSTEDGDLIAHWNSLTITELKKQAQQVFEATTNQQKATFNQLPDGIYYGL----AV 123
 Q E +TI E++ + + A + + G F +L + Y L V
 Sbjct: 1206 QQKVDE-----FVTIAEMETDEEGIVFAGSLEPGDYQFVELNAPVGYKLDDETPV 1256

Query: 124 KAGEKNRNVSAFLVDLSEDKVIYPKIIWSTGELDLLKVGVDGDTKKPLAGVVFELYEKNG 183
 E++R + ++L ++ + P G + L+KV D D L G F L + G
 Sbjct: 1257 FTVEEDRTET---IELQKENHLIP-----GSVQLVKVDAD-DAANTLEGAEFTLLDGE 1306

Query: 184 RTPIRVKNGVHSQDIDAAKHLETDSSGHIRISGLIHGDYVLKEIETQSGYQIGQAETAVT 243
 V+ G L TD +G + ++ L G+Y E + +GY++ T
 Sbjct: 1307 NV---VQEG-----LTTDENGQVVVTDLKPGEYQFVETKAPAGYELEATPIGFT 1352

Query: 244 IEKS--KTVTVTIENKKVP 260
 IE++ + TV +EN +P
 Sbjct: 1353 IERNQQEVATVAVENHLIP 1371

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3920 (GBS52) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 4; MW 30.5kDa).

GBS52-His was purified as shown in Figure 192, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1274

A DNA sequence (GBSx1351) was identified in *S.agalactiae* <SEQ ID 3921> which encodes the amino acid sequence <SEQ ID 3922>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.26 Transmembrane 554 - 570 (551 - 575)
 INTEGRAL Likelihood = -0.16 Transmembrane 34 - 50 (34 - 50)

----- Final Results -----
 bacterial membrane --- Certainty=0.3506(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8779> which encodes amino acid sequence <SEQ ID 8780> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0
 McG: Discrim Score: -5.81
 GvH: Signal Score (-7.5): -1.92
 Possible site: 37
 >>> Seems to have a cleavable N-terminal signal sequence
 ALOM program count: 2 value: -6.26 threshold: 0.0

-1411-

INTEGRAL Likelihood = -6.26 Transmembrane 527 - 543 (524 - 548)
 PERIPHERAL Likelihood = 5.36 194
 modified ALOM score: 1.75

5 *** Reasoning Step: 3

----- Final Results -----

10 bacterial membrane --- Certainty=0.3506(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 521-525

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAA57459 GB:X81869 orf2 [Lactobacillus leichmannii]
 Identities = 140/505 (27%), Positives = 220/505 (42%), Gaps = 94/505 (18%)

Query: 102 GEVISNYAKLGDNVKGLQGVQFKRYKVKTDI-----SVDELKKLTTVEAADAKVGITILEE 156
 GE+++++ G L GV FK Y V S D + T +DAK L

20 Sbjct: 58 GBIMNDPGGTG-----LNGVTFKAYNVTDHYLSLRKSGDSAQDAVTAIQSDAKSDNLPS 112

Query: 157 --GVSLPQKTNAQGLVVDAL-----DSKSNVR-YLYVEDLKNSPSNITKAYAVPFV 204
 G ++ +T A D + DS N + YL+VE +SP+++T+ A P V

25 Sbjct: 113 YAGSAIATETTATSKGEDGIAAFDNLNLKDSGNYQTYLFFVET--DSPTDVTQQ-AAPIV 169

Query: 205 LELPVANSTGTGFLS-EINIYPKNVVTDEPKTDKDVKKLGQDDAGYTI-----G 252
 L +P+ ++ T ++ +I IYPKNV + P T KD+ + + D T+ G

30 Sbjct: 170 LTMPIYKTSDTSAINHDIQIYPKNVKST-PIT-KDLDEASKKDLAVTLPDGSTIYNAQYG 227

Query: 253 EEFKWFCLKSTIPANLGDYEFKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVD 312
 + F + + +P N+ D + F + DK G+ + + L + YT+++

35 Sbjct: 228 KSPGYNITVNVPNWIKDKDTFNVVDKPDGTGI---DIDASTVSIDGLTKSTDYTVNK---- 280

Query: 313 NQNTLKITFKPEKFKEIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVL 372
 N ++ FK + L G +L I +T+ A

35 Sbjct: 281 KDNQYQVVFKTTS--AAVQALAGKSLT-----ITYKATLTNNATP 318

Query: 373 GKAIENFELQYDHTPDKADNPKNPNPPRKPEVHTGGKRFVKKDSTETQTLGGAEFDLA 432
 KAI NT L + + S P P ++TGG +FVKDS +TL GAEF L+

40 Sbjct: 319 DKAIGNATLSIGNGTNIT-----STPANGPRIYTGGAQFVKKDSQSNKTLGAEFQLVK 373

Query: 433 --SDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIKGLAYAVDANAEGT 490
 S+G V + + N A EA T S +G +KGL+Y ++ +

45 Sbjct: 374 VDSNGNIVSYATQASDGSYTWNDSATEATT-----YTSDANGLVALKGLSY---SDKLDS 425

Query: 491 AVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPN 550
 +Y L E +AP+GY D ++F+++Q S+ D+ TI N K +P+

50 Sbjct: 426 GESYALLEIQAPDGYAKLDSPVKFSITQGSF-----GDSNKITIDNTKEGLLPS 474

Query: 551 TGGIGTAIFVAIGAAMFAVKGMMK 575
 TGG G IF+AIG +M A G K

50 Sbjct: 475 TGGKGIYIFLAIGIVIMIVAFGGYK 499

No corresponding DNA sequence was identified in *S.pyogenes*.

55 SEQ ID 8780 (GBS80) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 6; MW 56.8kDa).

The GBS80-His fusion product was purified (Figure 104A; see also Figure 194, lane 5) and used to immunise mice (lane 1+2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 104B), FACS (Figure 104C), and in the *in vivo* passive protection assay (Table III). These tests confirm
 60 that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1275

A DNA sequence (GBSx1352) was identified in *S.agalactiae* <SEQ ID 3923> which encodes the amino acid sequence <SEQ ID 3924>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4043(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1276

A DNA sequence (GBSx1353) was identified in *S.agalactiae* <SEQ ID 3925> which encodes the amino acid sequence <SEQ ID 3926>. This protein is predicted to be MsmR. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.01 Transmembrane 75 - 91 ( 75 - 92)

----- Final Results -----
bacterial membrane --- Certainty=0.1404(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9679> which encodes amino acid sequence <SEQ ID 9680> was also identified.

SEQ ID 3926 (GBS360) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 9; MW 74kDa).

GBS360-GST was purified as shown in Figure 216, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1277

A DNA sequence (GBSx1354) was identified in *S.agalactiae* <SEQ ID 3927> which encodes the amino acid sequence <SEQ ID 3928>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1762(Affirmative) < succ>
```

-1413-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3929> which encodes the amino acid sequence <SEQ ID 3930>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1640 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 93/98 (94%), Positives = 96/98 (97%)

Query: 1 MDKIIKSISASGAFRSYVLDSTETVKLAQEKHHTLSSSTVALGRTLIANQILAANQKGDS 60
 MDKIIKSI+ SGAFR+YVLDSTETV LAQEKH+TLSSSTVALGRTLIANQILAANQKGDS
 Sbjct: 1 MDKIIKSIAQSGAFRAYVLDSTETVALAQEKHNTLSSSTVALGRTLIANQILAANQKGDS 60

Query: 61 KITVKVIGDSSFGHIISVADTKGHVKGYYIQTGVDIKK 98
 KITVKVIGDSSFGHIISVADTKGHVKGYYIQTGVDIKK
 Sbjct: 61 KITVKVIGDSSFGHIISVADTKGHVKGYYIQTGVDIKK 98

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1278

A DNA sequence (GBSx1355) was identified in *S.agalactiae* <SEQ ID 3931> which encodes the amino acid sequence <SEQ ID 3932>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC98436 GB:L29324 unknown [Streptococcus pneumoniae]
 Identities = 34/48 (70%), Positives = 39/48 (80%)
 Query: 1 MQEVLIIARENHQVTHEHVSILLTCVQELIVEVNQTQPLSREFREKYM 48
 + EV IIA+ NHQVTHEHVSILLTC+QELI EV +T PLS +F KYM
 Sbjct: 70 VHEVFIIAKTNHQVTHEHVSILLTCIQELIKEVEKTGPLESDFCNKYM 117

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1279

A DNA sequence (GBSx1356) was identified in *S.agalactiae* <SEQ ID 3933> which encodes the amino acid sequence <SEQ ID 3934>. This protein is predicted to be TnpA (orfB). Analysis of this protein sequence reveals the following:

-1414-

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.5248(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9907> which encodes amino acid sequence <SEQ ID 9908>
10 was also identified. A further related GBS nucleic acid sequence <SEQ ID 9677> which encodes amino acid
sequence <SEQ ID 9678> was also identified. A further related GBS nucleic acid sequence <SEQ ID
10911> which encodes amino acid sequence <SEQ ID 10912> was also identified.

There is homology to SEQ ID 1336.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
15 vaccines or diagnostics.

Example 1280

A DNA sequence (GBSx1357) was identified in *S.agalactiae* <SEQ ID 3935> which encodes the amino
acid sequence <SEQ ID 3936>. Analysis of this protein sequence reveals the following:

Possible site: 45
20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.4489(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB64982 GB:U43834 Ydr540cp [Saccharomyces cerevisiae]
Identities = 93/171 (54%), Positives = 121/171 (70%), Gaps = 3/171 (1%)

30 Query: 1 MRVYENKEELKKEISKTFEKYIMEFNNIPENLKDKRIDEVDRTTPAANLSYQVGWTLNLVLK 60
 MR Y +K+ELK+EI K +EKY EF I E+ KD++++ VDRTP+ NLSYQ+GW NL+L+
Sbjct: 1 MREYTSKKELKKEIEKKYKDYDAEFETISESQKDEKVETVDRTPSENLSYQLGWNLLLE 60

35 Query: 61 WEEDERKGLQVKTPSDKFKWNQLGELYQWFTDTYAHLSLQELKAKLNENINSIYAMIDLL 120
 WE E G V+TP+ +KWN LG LYQ F Y S++E +AKL E +N +Y I L
Sbjct: 61 WEAKEIAGYNVETPAPGYKWNNGGLYQSFYKKYGIYSIKEQRAKLREAVNEVYKWISTL 120

40 Query: 121 SEEELFEAHMRKWADEATKTATWEVYKFIHVNTVAPFGTFRFKIRKWKKIV 171
 S++ELF+A RKW AT A W VYK+IH+NTVAPF FR KIRKWK++V
Sbjct: 121 SDDELFGAGNRKW--ATTKAMWPVYKWIHINTVAPFTNFRGKIRKWKRLV 168

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
45 vaccines or diagnostics.

Example 1281

A DNA sequence (GBSx1358) was identified in *S.agalactiae* <SEQ ID 3937> which encodes the amino
acid sequence <SEQ ID 3938>. Analysis of this protein sequence reveals the following:

Possible site: 28
50 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.45 Transmembrane 10 - 26 (2 - 26)

-1415-

----- Final Results -----

bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8781> which encodes amino acid sequence <SEQ ID 8782> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 8.80
 GvH: Signal Score (-7.5): -3.94
 Possible site: 28
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -3.45 threshold: 0.0
 INTEGRAL Likelihood = -3.45 Transmembrane 7 - 23 (2 - 26)
 PERIPHERAL Likelihood = 10.40 69
 modified ALOM score: 1.19

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear)

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA68889 GB:Y07615 acid phosphatase [Haemophilus influenzae]
 Identities = 112/245 (45%), Positives = 148/245 (59%), Gaps = 10/245 (4%)

Query: 5 MKKVLVSSLLVLGITITLQTVVEAKGPKVAYTQEGMTALSDTNKDKVTTISIDEIQKSLE 64
 MK V+ S++ L +T V G YTQ G A + + IS+D+I++SLE
 Sbjct: 1 MKNVMKLSVIAL---LTAAAVPAMAGKTEPYTQSGTNAREMLQEQAIHWISVDQIKQSLE 57

Query: 65 GKKPITVSFDDITLLFSSQYFQYQKEYVTPGSFDFLHKQKFWDLVAKRGDQDSIPKEYA 124
 GK PI VSFDDIT+LFSS F +G++ +PG D+L Q FW+ V D+ SIPK+ A
 Sbjct: 58 GKAPINVSFDDITVLFSSPCFYHGQKQKFSFGKHDYLNQDFWNEVNAGCDKYSIPKQIA 117

Query: 125 KKLIAMHQKRGDKIVFITGRTRGSMYKEGEVDKTAALAKDFKLDKPIAVNYTGDKPKKP 184
 LI MHQ RGD++ F TGRT G+VD L K F + V + G + ++
 Sbjct: 118 IDLINMHQARGDQVYFFTGRT-----AGKVDGVTPILEKTFNKNMHPVEFMGSR-ERT 170

Query: 185 YKYDKSYYIKKYGSDIHYGSDDDDIHAAREAGARPILRAPNSTNLPLPEAGGYGEEVL 244
 KY+K+ I + IHYGSDDD+ AA+EAG R IR++RA NST P+P GGYGEEVL
 Sbjct: 171 TKYNKTPAIISHKVSIIHYGSDDDVLAAKEAGVGRGIRLMRAANSTYQPMPTLGGYGEEVL 230

Query: 245 ENSAY 249

NS+Y

Sbjct: 231 INSSY 235

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3939> which encodes the amino acid sequence <SEQ ID 3940>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.98 Transmembrane 6 - 22 (4 - 25)

----- Final Results -----

bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA68889 GB:Y07615 acid phosphatase [Haemophilus influenzae]

-1416-

Identities = 105/237 (44%), Positives = 141/237 (59%), Gaps = 10/237 (4%)

Query: 9 LFTVSFCGIIALPVEASGPKVPYTQEGITA--ISNQATVKLISIADIASSLEGQKPITVS 66
 L ++ A+P A G PYTQ G A + + + IS+ I SLEG+ PI VS
 5 Sbjet: 7 LSVIALLTAAAVPAMA-GKTEPYTQSGTNAREMLQEQAIHWISVDQIKQSLEGKAPINVS 65

Query: 67 FDIDDTLLFTSQYFQYGKEYITPGSFDFLHKQKFWDLVAKRGDQDSIPKEYAKQLIAMHQ 126
 FDIDDT+LF+S F +G++ +PG D+L Q FW+ V D+ SIPK+ A LI MHQ
 10 Sbjet: 66 FDIDDTVLFSSPCFYHGQKQKFSFGKHDYLNQDFWNEVNAGCDKYSIPKQIAIDLINMHQ 125

Query: 127 KRGDKIVFITGRTRGSMYKKGEIDKTAKSLAKDFKLDKPIAINYTGDKAVKPYQYDKTTY 186
 RGD++ F TGRT G++D L K F + + + G + + +Y+KT
 15 Sbjet: 126 ARGDQVYFFFTGRT-----AGKVDGVTPILEKTFNIKNMHPVEFMGSRE-RTTKYNKTPA 178

Query: 187 IKKNGSQIHYGDSDEDINAAKEAGARPILRAPNSTNLPLPKAGGYGEEVLNSAY 243
 I + IHYGSD+D+ AAKEAG R IR++RA NST P+P GGYGEEVL NS+Y
 15 Sbjet: 179 IISHKVISIHYGSDDDVLAKEAGVRGIRLMRAANSTYQPMPTLGGYGEEVLINSSY 235

An alignment of the GAS and GBS proteins is shown below.

Identities = 196/245 (80%), Positives = 216/245 (88%), Gaps = 2/245 (0%)

Query: 5 MKKVLVSSLLVLGITITLQTVVEAKGPKVAYTQEGMTALSDTNKDKVTTISIDEIQKSLE 64
 MKK S L + + VEA GPKV YTQEG+TA+S N+ V ISI +I SLE
 20 Sbjet: 1 MKKEFTSILFTVSFCGIIALPVEASGPKVPYTQEGITAIS--NQATVKLISIADIASSLE 58

Query: 65 GKKPITVSFDIDDTLLFSSQYFQYGKEYVTPGSFDFLHKQKFWDLVAKRGDQDSIPKEYA 124
 G+KPIITVSFDIDDTLLF+SQYFQYGKEY+TPGSFDFLHKQKFWDLVAKRGDQDSIPKEYA
 25 Sbjet: 59 GQKPITVSFDIDDTLLFTSQYFQYGKEYITPGSFDFLHKQKFWDLVAKRGDQDSIPKEYA 118

Query: 125 KKLIAHQKRGDKIVFITGRTRGSMYKEGEVDKTAKALAKDFKLDKPIAVNYTGDKPKKP 184
 K+LIAHQKRGDKIVFITGRTRGSMYK+GE+DKTAK+LAKDFKLDKPIA+NYTGDK KP
 30 Sbjet: 119 KQLIAHQKRGDKIVFITGRTRGSMYKKGEIDKTAKSLAKDFKLDKPIAINYTGDKAVKP 178

Query: 185 YKYDKSYIYIKKYGSDIHYGSDDDIHAAREAGARPILRAPNSTNLPLPEAGGYGEEVL 244
 Y+YDK+YYIKK GS IHYGSD+DI+AA+EAGARPILRAPNSTNLPLP+AGGYGEEVL
 35 Sbjet: 179 YQYDKTYIYIKKNGSQIHYGSDDEDINAKEAGARPILRAPNSTNLPLPKAGGYGEEVL 238

Query: 245 ENSAY 249
 ENSAY
 40 Sbjet: 239 ENSAY 243

SEQ ID 8782 (GBS100) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 5; MW 28kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 2; MW 53kDa).

The GBS100-GST fusion product was purified (Figure 106A; see also Figure 197, lane 4) and used to immunise mice (lane 1 product; 9.9µg/mouse). The resulting antiserum was used for Western blot (Figure 106B), FACS (Figure 106C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1282

A DNA sequence (GBSx1359) was identified in *S.agalactiae* <SEQ ID 3941> which encodes the amino acid sequence <SEQ ID 3942>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

-1417-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3288(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 1283

A DNA sequence (GBSx1360) was identified in *S.agalactiae* <SEQ ID 3943> which encodes the amino acid sequence <SEQ ID 3944>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4004(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9675> which encodes amino acid sequence <SEQ ID 9676> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04406 GB:AP001509 RNA methyltransferase [Bacillus halodurans]
 Identities = 198/452 (43%), Positives = 300/452 (65%)

25

Query: 12 KRKIMLHKNDIETEISDISHEGMGIKVDGFFVFFVENALPGEIIMRVLKLKRIGYGK 71
 K++ ++KND++E I D++H+G G+AKVDG+ F+ ALPGE +K +V+K++K G+G+
 Sbjct: 3 KQAPVKNKNDVVEVTIEDLTHDGAGVAKVDGYALFIPKALPGERLKAKVVKVKKGYGFGR 62

30

Query: 72 VEEYLTTSPHRNEGLDYTLRTGIADLGHLYEQQLLFKQKQVADNLYKIAHISDVLVEP 131
 V + SP R E + + G L H++Y+ QL +KQKQV D L +I I+ V V P
 Sbjct: 63 VLNMIASPDREAPCPVFNQCGGCQLQHMSYDAQLRYKQKQVQDVLERIGKITAVTVRP 122

35

Query: 132 TLGMTIPLAYRNKAQVPVRRVDGQLETGFFRKNSHTLVSIEDYLIQKEIDALINFTRDL 191
 T+GM P YRNKAQVPV +G L GF+++ SH ++ +++ +IQ +E D +I ++L
 Sbjct: 123 TIGMNEPWRYRNKAQVPVGEREGGLIAGFYQERSHRIIDMDECMIQHEENDKVIQVKEL 182

40

Query: 192 LRKFDVKPYDEEQSGLIRNLVVRGHTGQMLVLVTRPKIFRIDQMIEKLVSAPPSV 251
 R+ ++ YDEE+ G +R++V R G TG++M+VL+T ++ +IE++ A P V
 Sbjct: 183 ARELGIRGYDEEKHRTLRHVARYGKNTGEIMVVLITRGEELPHKKTIERIHKAIPHV 242

45

Query: 252 VSIMQNINDRNSNVIFGKEFRTLYGSDTIEDQMLGNTYAISAQSFYQVNTMAEKLYOKA 311
 SI+QN+N + +NVIFG + + L+G + I D + +AISA+SFYQVN E + LY +A
 Sbjct: 243 KSIQVQNVNPKRTNVIFGDKTKVLWGEEFYIYDTIGDIKFAISARSFYQVNPQTKVLVDQA 302

50

Query: 312 IDFSDLNSEDIVIDAYSIGITIGLSVAKQVKHVGVEVVEKAVSDAKENATRNGITNSTY 371
 ++F++L + VIDAY GIGTI L +A+Q KHVGVE+V +A+SDAK NA NG N +
 Sbjct: 303 LEFANLTGSETVIDAYCGIGTISLFLAQAKHVGVEIVPEAISDAKRNRNLNGFANVQF 362

Query: 372 VADSAENAMAKWLKEGIKPTVIMVDPKRGLTESFVYSAAQTKADKITYISCNSATMARD 431
 AE M W +G++ VI+VDPPRKG E+ + + K D++ Y+SCN AT+ARD
 Sbjct: 363 AVGDAEKVMPWYQAQGVRAVDIVVDPKRKGCDEALLKTIILNMKPDVVVYVSCNPATLARD 422

55

Query: 432 IKLFEELGYHLVKIQPVDLFPMTTHHVECVALL 463
 +++ E+ GY +QPVD+FP T H+E VA+L
 Sbjct: 423 LRVLEDGGYETKDVQPVDMFPWTTHIESVAVL 454

-1418-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3945> which encodes the amino acid sequence <SEQ ID 3946>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1262(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 332/454 (73%), Positives = 387/454 (85%)

Query: 12 KRKIMLHKNDIIEETEISDISHEGMGIKVDGFFVFFVENALPGEIIMRVLKLKRKIGYGK 71

KR ML KNDII+ ISD+SHEG G+AK DGFVFFV+NALP E+I MRVLK+ K G+GK

Sbjct: 8 KRIRMLKKNDDIQVAISDLSEHGAGVAKHDGFFVFFVDNALPEEVIDMRVLKVNKNSGFGK 67

Query: 72 VEEYLTTSPHRNEGLDYTYLRTGIADLGHLYEQQLLFKQKQVADNLYKIAHISDVLVEP 131

VE Y S RN ++ TYLRTGIADLGHLYEQ QL FK+KQV D+LYKIA ISDV VE

Sbjct: 68 VEAYHYLSSARNADVNLTYLRTGIADLGHLYEQQLTFKKKQVQDSLYKIAGISDVTVES 127

Query: 132 TLGMTIPLAYRNKAQVPVRRVDGQLETGFFRKNSHTLVSIEDYLIQEKEIDALINFTRDL 191

T+GMT PLAYRNKAQVPVRRV+GQLETGFFRK+SH L+ I DY IQ+KEID LINFTRDL

Sbjct: 128 TIGMTEPLAYRNKAQVPVRRVNGQLETGFFRKSHDLIPISDYIYQKEIDRLINFTRDL 187

Query: 192 LRKFDVKPYDEEQSGLIRNLVRRGHYTGQMLVLVTTTRPKIFRIDQMIEKLVSAPPSV 251

LR+FD+KPYDE +Q+GL+RN+VRRGHY+G++MLVLVTTTRPK+FR+DQ+IEK+V AFP+V

Sbjct: 188 LRRFDIKPYDETEQTGLLRNIVRRGHYSGEMMLVLVTTTRPKVFRVDQVIEKIVEAPPAV 247

Query: 252 VSIMQNINDRNSNVIFGKEFRTLYGSDTIEDQMLGNTYAISAQSFYQVNTMAEKLYQKA 311

VSI+QNIND+N+N IFGK+F+TYG DTI D MLGN YAISAQSFYQVNT MAEKLYQ A

Sbjct: 248 VSIIQNINDKNTNAIFGKDFKTLYGKDTITDSMLGNYYAISAQSFYQVNTVMAEKLYQTA 307

Query: 312 IDFSDLNSEDIVIDAYSGIGTIGLSVAKQVKHVGVEVVEKAVSDAKENATRNGITNSTY 371

I FSDL+ +DIVIDAYSGIGTIGLS AK VK VYGVEV+E AV DA++NA NGITN+ +

Sbjct: 308 IAFSDLKDDIVIDAYSGIGTIGLSFAKTIVKAVYGVEVIEAAVRDAQQNAALNGITNAYF 367

Query: 372 VADSAENAMAKWLKEGIKPTVIMVDPPrKGLTESFVYSAAQTKADKITYISCNSATMARD 431

VAD+AE+AMA W K+GIKP+VI+VDPPrKGLTESF+ ++ KITY+SCN ATMARD

Sbjct: 368 VADTAEHAMATWAKDGIKPSVILVDPPrKGLTESFIQASVAMGPQKITYVSCNPATMARD 427

Query: 432 IKLFEELGYHLVKIQPVDLFPMTTHVECVALLVK 465

IK ++ELGY L K+QPVDLFP THVECV LL+K

Sbjct: 428 IKRYQELGYKLAKVQPDVLPQTHVECVVLLIK 461

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1284

A DNA sequence (GBSx1361) was identified in *S.agalactiae* <SEQ ID 3947> which encodes the amino acid sequence <SEQ ID 3948>. This protein is predicted to be PSR protein. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-12.15 Transmembrane 135 - 151 (127 - 155)

----- Final Results -----

bacterial membrane --- Certainty=0.5861(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1419-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB76822 GB:AJ276232 PSR protein [Enterococcus faecalis]
Identities = 143/409 (34%), Positives = 206/409 (49%), Gaps = 56/409 (13%)

5 Query: 48 QRRTESPP--TNSYYEPEYSDSYQQDDDFYSEPOLTSQGLPIYQEERAPKKKQRRARKEK 105
+ R E P S E Y DSY +D T G ++ P+ KK + K+K
Sbjct: 31 EHREEEPEELAESLQEPVYEDSYTEDSRRSERRHQTDSGGG-NGSDQPPRGKKDKKPKKK 89

10 Query: 106 QRVKVMAPFPPKAITPPRKKKKFKGFLKFIGIILLIVLSGMVFMFVKGMRDVNNKSHYS 165
RKK K K F K++ I+L+++ + MF+KG + S
Sbjct: 90 -----RKKSKTKRFFKWLVIILLILLFAYSTMFLKKGKSAAEHDDS-LP 131

15 Query: 166 PAIEDFKGKDAVDGT-NILILGSDKRVSERSTDARTDTIMVANVGKDNKVKMVSFMRD 224
+E F G + +G NILILGSD R + R DTIMV + K K++SFMRD
Sbjct: 132 QEKVETFNQVSSNGAKNILILGSDTRGEDAG---RADTIMVLQNLNGPSKKPKLISFMRD 188

20 Query: 225 LLVNIPNYSTEGYDMKLNASFNLGEQDNHKGAEYVRQTLKNHFDIDIKYYVMVDFETFA 284
V+IP G K+NA++ G GAE VR+TLK +F++D KYY VDF++F
Sbjct: 189 TFVDIP----GVGPNKINAAYAG-----GAELVRETLKQNFNLDTKYYAKVDFQSFE 237

25 Query: 285 DAIDTLFPNGVKINAKFGLVGGQSADSVKVPDDLRLMKNQVVPVPSQKIKVGIQYMDGRTLIN 344
+D++FP GVKI+A+ L + D V I+ G Q MDG LL
Sbjct: 238 KIVDSMFPGVKIDAESL-----NLDGVD-----IEKGQQVMDGHVLLQ 277

30 Query: 345 YARFRKDDGDFGRTRQQQVMRAIVSQIKDPRRLFTGSAAIGKAYALTSSNLSYSFVLIT 404
YARFR D++GDFGR +RQQQVM A++SQ+K+P L ++GK S+++ SF+LT
Sbjct: 278 YARFRMDEEGDFGRVRRQQQVMSAVMSQMKNPMTLLRTPESLGKLVGYMSTDVPVSFMLT 337

Query: 405 DGIPILSDAKNGIKQMTIPREGDWVDDYDQYGGQGLTIDFAKYKKILKK 453
+G +L K G++ +++P W Y G L +D K ++K
Sbjct: 338 NGPSLLIKGKTGVESLSVPVPSWVNFGESSYAGSILEVDEQKNADAIEK 386

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3949> which encodes the amino acid sequence <SEQ ID 3950>. Analysis of this protein sequence reveals the following:

35 Possible site: 49

>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -7.96 Transmembrane 159 - 175 (152 - 180)

40 ----- Final Results -----
bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the databases:

>GP:CAB76822 GB:AJ276232 PSR protein [Enterococcus faecalis]
Identities = 140/345 (40%), Positives = 195/345 (55%), Gaps = 41/345 (11%)

50 Query: 140 PRSQK---RKHKKKGCMKWFFNILGLLLMTVLMGLGIMFAKGVFDISTNKANYKPAVSQ 195
PR +K +K +KK K FF L +LL+ + +MF KG + + + V +
Sbjct: 78 PRGKKDKKPKKKRKKSKTKRFFKWLVIILLILLFAYSTMFLKKGKSAAEHDDSLPQEKV-E 136

55 Query: 196 AFDGQETQDGT-NILILGSDQRTQGSTDARTDTIMVVNVGNHAKKIKMVSFMRDTLINI 254
F+G ++ +G NILILGSD T+G R DTIMV+ + +KK K++SFMRDT ++I
Sbjct: 137 TFNGVKSSNGAKNILILGSD---TRGEDAGRADTIMVLQNLNGPSKKPKLISFMRDTPVDI 193

60 Query: 255 PGYSYNDNSYDLKLNSAFNLGEQEDHHGAEYVRRALKHNFIDIDIKYYVMVDFETFAEAI 314
PG N K+N+A+ G GAE VR LK NF++D KYY VDF++F + +D
Sbjct: 194 PGVGN-----KINAAYAG-----GAELVRETLKQNFNLDTKYYAKVDFQSFEKIVD 241

Query: 315 TLFNGVKIDAKFATVGGVAVDSVEVPDDLRLMKNQVVPNQTIEVGEQRMGRTLLNYARF 374
++FP GVKIDA+ + + +D V+ IE G+Q MDG LL YARF
Sbjct: 242 SMFPGVKIDAESL-----LNLDGVD-----IEKGQQVMDGHVLLQYARF 281

-1420-

Query: 375 RKDEEGDFGR TVRQQQVMSAVMSQIKDPTKLTGSAAGKIYALTSTNVSFPPVVKNGVS 434
 R D+EGDFGR RQQQVMSAVMSQ+K+P L ++GK+ ST+V F++ NG S
 Sbjct: 282 RMDEEGDFGRVRRQQQVMSAVMSQMKNPMTLLRTPESLGKLVGYMSTDPVPSFMLTNGPS 341

5 Query: 435 VLGSCKNGVEHVTIPENGDWVDEYDMYGGQALYIDFDKYQKTLAK 479
 +L GK GVE +++P W Y G L +D K + K
 Sbjct: 342 LLIKGTGVESLSVPVPDSWNFGESSYAGSILEVDEQKNADAIEK 386

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 273/486 (56%), Positives = 340/486 (69%), Gaps = 32/486 (6%)

Query: 1 MSRNNGYQLNHHEELRYNYLLKNIHYLNEREKMEFQYLHYKKTAVRPQRRTESPTNSYY 60
 M++ G L+HHEELRY YLL+N+ YL+E EK EF +L K R ++ S
 Sbjct: 1 MTKYPMGGLSHHEELRYFYLLRNLSYLSENEKKEFAFLKSKLEIGRAYAPSKQHYRKSQR 60

15 Query: 61 EEPY-SDSY-----QDDDFYSEPLTSQGLPIYQEERAPKKKQARKEKQRVKV 110
 +EPY D YY +DDD + GLPIY +E KK K R +
 Sbjct: 61 QEPYFEDDYNDYSPNDLLEDDDVNHDSSFVYPYGLPIYPKEDRYLNKKT---KLTARRPI 117

20 Query: 111 MAPFP-----PKAITPPRKKKK-FKGFLKFIGIILLJVLSGVMFMFVK 152
 AP P P++ KKK K F +G++L+ VL G+ MF K
 Sbjct: 118 DAPQPIDEDDAFLTESVARCALPRSQKRKHKKKGCMKWFNLLGLLLMTVLMGLGLMFAK 177

25 Query: 153 GMRDVNNGKSHYSPAIEDFKGKDAVDGTNILLGSDKRVSESTDARTDTIMVANVGNK 212
 G+ D++ K++Y PA+ + F G++ DGTNILLGSD+RV++ STDARTDTIMV NVGN
 Sbjct: 178 GVFDISTNKANYKPAVSQAQFDGQETQDGTNILLGSDQRVVTQGSTDARTDTIMVNVGNH 237

30 Query: 213 DNKVKMVSFMRDLLVNIPNYS-TEGYDMKLNASFNLGEQDNHKGAEYVRQTLKNHFDID 271
 K+KMVSFMRD L+NIP YS + YD+KLN++FNLGEQ++H GAHYVR+ LK++FDID
 Sbjct: 238 AKKIKMVSFMRDTLINIPGYSYNDNSYDLKLNASFNLGEQEDHHGAHYVRRALKHNFDID 297

35 Query: 272 IKYYVMVDFETFAIDAIDTLFPNGVKINAKFGLVGGQSADSVKVPDDLRLMKNGVVPSQKIK 331
 IKYYVMVDFETFA+AIIDTLFPNGVKI+AKF VGG + DSV+VPDDLRLMKNGVVP+Q I+
 Sbjct: 298 IKYYVMVDFETFAIDAIDTLFPNGVKIDAKFATVGGVAVDSEVPDDLRLMKNGVVVPNTIE 357

40 Query: 332 VGIQYMDGRTLLNYARFRKDDGDFGRTRQQQVMRAIVSQIKDPRRLFTGSAAGKAYA 391
 VG Q MDGRTLLNYARFRKDD+GDFGRTRQQQVM A++SQIKDP +LFTGSAAGK YA
 Sbjct: 358 VGEQRMDGRTLLNYARFRKDDGDFGRTRQQQVMSAVMSQIKDPTKLTGSAAGKIYA 417

45 Query: 392 LTSSNLSYSFVLTGIPILSDAKNGIKQMTIPREGDWDDYDQYGGQGLTIDFAKYKKIL 451
 LTS+N+S+ FV+ +G+ +L KNG++ +TIP GDWVD+YD YGGQ L IDF KY+K L
 Sbjct: 418 LTSTNVSFPPVVKNGVSVLGSCKNGVEHVTIPENGDWVDEYDMYGGQALYIDFDKYQKTL 477

Query: 452 KKMGLR 457
 K+GLR
 Sbjct: 478 AKLGLR 483

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 1285

A DNA sequence (GBSx1362) was identified in *S.agalactiae* <SEQ ID 3951> which encodes the amino acid sequence <SEQ ID 3952>. This protein is predicted to be shikimate kinase (aroK). Analysis of this protein sequence reveals the following:

55 Possible site: 17
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 60 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1421-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA55181 GB:X78413 shikimate kinase [Lactococcus lactis]
Identities = 65/164 (39%), Positives = 98/164 (59%), Gaps = 8/164 (4%)

5  Query: 1  MPKVLLGFMGVGKTSVANCLENEVIDMDSLIEKHIGMSISRFFTEEGEASFRALQSFLN 60
      M +L+GFMG GK++VA L E D+D LIE+ I M I+ FF GEA FR +E++
      Sbjct: 1  MSIILIGFMGAGKSTVAKLLAEFTDLDKLIEEEIEMPIATFFELFGEADFRKIENEVFE 60

      Query: 61  ELLKKKNEGLVIASGGGIVLLEENRRLTLNRHNNIL-LTGSFEVLYHRIKKDEKNRRPL 119
      ++K ++IA+GGGI+ E + L L+R + ++ LT F+ L+ RI D +N RP
      Sbjct: 61  LAVQK---DIIATGGGII--ENPKNLNVLDRASRVVFLTADFDTLWKRI SMDWQNVRP- 114

      Query: 120  FLNHSKEEFYDIYQKRMLLYSGLSDMIIDTDLTPQKIATVIGE 163
      L KE +++KRM YS ++D+ ID +P++IA I E
      Sbjct: 115 -LAQDKAAQLLFKRMKDYSLVADLTIDVTDKSPEQIAEQIRE 157
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3953> which encodes the amino acid sequence <SEQ ID 3954>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA55181 GB:X78413 shikimate kinase [Lactococcus lactis]
Identities = 63/160 (39%), Positives = 97/160 (60%), Gaps = 5/160 (3%)

30  Query: 1  MTKVLLGFMGVGKTTVSKHLSMHCKDMDAII EAKIGMSIAAFFEQHGEIAFRTIESQVLK 60
      M+ +L+GFMG GK+TV+K L+ D+D +IE +I M IA FFE GE FR IE++V +
      Sbjct: 1  MSIILIGFMGAGKSTVAKLLAEFTDLDKLIEEEIEMPIATFFELFGEADFRKIENEVFE 60

      Query: 61  DLLFANDNSIIVTGGGVVVLQENRQLLRKNHQNILLVASFETLYQRLKHKKSQRPLFL 120
      L + II TGGG++ +N +L + + L A F+TL++R+ D ++ RP L
      Sbjct: 61  --LAVQKDIIATGGGIIENPKNLNVLDLDR-ASRVVFLTADFDTLWKRI SMDWQNVRP--L 115

      Query: 121  KYSKEAFYEFYQORMVFEGLSDLVIRVDHRTPEEVANII 160
      KEA +++RM Y ++DL I V ++PE++A I
      Sbjct: 116 AQDKAAQLLFKRMKDYSLVADLTIDVTDKSPEQIAEQI 155
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 88/161 (54%), Positives = 120/161 (73%), Gaps = 1/161 (0%)

45  Query: 1  MPKVLLGFMGVGKTSVANCLENEVIDMDSLIEKHIGMSISRFFTEEGEASFRALQSFLN 60
      M KVLGFMGVGKT+V+ L DMD++IE IGMSI+ FF + GE +FR +ESQ L
      Sbjct: 1  MTKVLLGFMGVGKTTVSKHLSMHCKDMDAII EAKIGMSIAAFFEQHGEIAFRTIESQVLK 60

      Query: 61  ELLKKKNEGLVIASGGGIVLLEENRRLTLNRHNNILLTGSFEVLYHRIKKDEKNRRPLF 120
      +LL N+ +I +GGG+V+L+ENR+LL N +NIL SFE LY R+K D+K++RPLF
      Sbjct: 61  DLLFA-NDNSIIVTGGGVVVLQENRQLLRKNHQNILLVASFETLYQRLKHKKSQRPLF 119

      Query: 121  LNHSKEEFYDIYQKRMLLYSGLSDMIIDTDLTPQKIATVI 161
      L +SKE FY+ YQ+RM+ Y GLSD++I D+ TP+++A +I
      Sbjct: 120 LKYSKEAFYEFYQORMVFEGLSDLVIRVDHRTPEEVANII 160
```

SEQ ID 3952 (GBS152) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 25 (lane 2; MW 20kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 2; MW 45.5kDa).

-1422-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1286

A DNA sequence (GBSx1363) was identified in *S.agalactiae* <SEQ ID 3955> which encodes the amino acid sequence <SEQ ID 3956>. This protein is predicted to be 3-phosphoshikimate 1-carboxyvinyltransferase (aroA). Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.81 Transmembrane 241 - 257 (240 - 257)

INTEGRAL Likelihood = -0.06 Transmembrane 390 - 406 (390 - 406)

----- Final Results -----

bacterial membrane --- Certainty=0.1723(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9673> which encodes amino acid sequence <SEQ ID 9674> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD45819 GB:AF169483 5-enolpyruvylshikimate-3-phosphate synthase

[Streptococcus pneumoniae]

Identities = 288/426 (67%), Positives = 347/426 (80%)

Query: 5 MKLLTNANTLKGITIRVPGDKSISHRAIIFGSISQGVTRIVDVLRGEDVLSTIEAFKQMGV 64

MKL TN L G IRVPGDKSISHR+IIFGS+++G T++ D+LRGEDVLST++ F+ +GV

Sbjct: 1 MKLKTNIIRHLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVFRDLGV 60

Query: 65 LIEDDGEIITTYGKGFAGLTQPNLLDMGNSGTSMLRIAGVLAGQEFVMTMGDNLSSKR 124

IED +IT+ G G AGL P N L+MGNSGTS+RLI+GVLAG +FEV M GD+SLSKR

Sbjct: 61 EIEDKDGIVITVQGVGMAGLKAPQNALNMGNSGTSIRLISGVLAGADFEVEMFGDDSLSKR 120

Query: 125 PMDRIALPLSKMGARISGVTNRDLPLKLQGTKKLKPIFYHLPVASAQVKSALIFAALQT 184

PMDR+ LPL KMG ISG T RDLPL+L+GTK L+PI Y LP+ASAQVKSAL+FAALQ

Sbjct: 121 PMDRVTLPLKMGVSIISGQTERDLPLRLKGTKNLRPIHYELPIASAQVKSALMFAALQA 180

Query: 185 KGESLIVEKEQTRNHTEDMIRQFGGHLDIKKEIRLNGGQSLVGDQDIRVPGDISSAFWI 244

KGES+I+EKE TRNHTEDM++QFGGHL + K+I + G Q L GQ + VPGDISSAFW+

Sbjct: 181 KGESVIIIEKEYTRNHTEDMLQQFGGHLSDVGKKITVQGPQKLTGQKVVPVPGDISSAFWL 240

Query: 245 VAGLIIPNSHIILENVGINETRTGILDVVSVMGGKIKLSSVDNQVKSATLTVDYSHLQAT 304

VAGLI PNS ++L+NVGINETRTGI+DV+ MGGK++++ +D KSATL V+ S L+ T

Sbjct: 241 VAGLIAPNSRLVLQNVGINETRTGIIDVIRAMGGKLEITEIDPVAKSATLIVESSDLKGT 300

Query: 305 HISGAMIPRLIDELPIIALLATQAQGTTVIADAQELKVKETDRIQVVESLQMGADITA 364

I GA+IPRLIDELPIIALLATQAQG TVI DA+ELKVKETDRIQVV ++L MGADIT

Sbjct: 301 EICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKVKETDRIQVVADALNSMGADITP 360

Query: 365 TADGMIIIRGNTPLHAASLDCHGDHRIGMMIAAALLVKEGEVDLSGEEAINTSYNPFLEH 424

TADGMII+G + LH A ++ GDHRIGMM AIAALLV +GEV+L EAINTSYP+F +

Sbjct: 361 TADGMIIKGSALHGAVNTFGDHRIGMMTATAALLVADGEVELDRAEAINTSYSPFFDD 420

Query: 425 LEGLVN 430

LE L++

Sbjct: 421 LESLIH 426

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3957> which encodes the amino acid sequence <SEQ ID 3958>. Analysis of this protein sequence reveals the following:

Possible site: 36

-1423-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.18 Transmembrane 240 - 256 (239 - 256)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAD45819 GB:AF169483 5-enolpyruvylshikimate-3-phosphate synthase
 [Streptococcus pneumoniae]
 Identities = 278/426 (65%), Positives = 346/426 (80%)

15 Query: 4 MKLRTNAGPLQGTIQVPGDKSISHRAVILGAVAKGETRVKGLLKGEDVLSTIQAFRNLGV 63
 MKL+TN L G I+VPGDKSISHR++I G++A+GET+V +L+GEDVLST+Q FR+LGV
 Sbjct: 1 MKLKTNIHRLHGIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVFRDLGV 60

20 Query: 64 RIEEKDDQLVIEGQGFQGLNAPCQTLNMGNSGTSMRLIAGLLAGQPFVSKMIGDESLSKR 123
 IE+KD + ++G G GL AP LNMGNSGTS+RLI+G+LAG F V+M GD+SLSKR
 Sbjct: 61 EIEDKDGVITVQGVGMAGLKAPQNALNMGNSGTSIRLISGVLAGADFEVEMFGDDSLSKR 120

25 Query: 124 PMDRIVYPLKQMGVEISGETDRQFPPLQLQGNRNLPITYTLPISSAQVKSAILLAALQA 183
 PMDR+ PLK+MGV ISG+T+R PPL+L+G +NL+PI Y LPI+SAQVKSAA+ AALQA
 Sbjct: 121 PMDRVTLPKMGVVISGQTERDLPLRLKGTKNLRPIHYELPIASAQVKSALMFAALQA 180

30 Query: 184 KGTTQVVEKEITRNHTEEMIQQFGRLIVDGKRITLVGPQQLTAQEITVPGDISSAAFVWL 243
 KG + ++EKE TRNHTE+M+QFGG L VDGK+IT+ GPQ+LT Q++ VPGDISSAAFVWL
 Sbjct: 181 KGESVIIKEYTRNHTEDMLQQFGGHLSDVGKKITVQGPQKLTGQKVVPVPGDISSAAFVWL 240

35 Query: 244 VAGLIIPGSELLKQNVGNPTRTGILEVVEKMGQAQIVYEDMNKKEQVTSIRVVYSNMKGT 303
 VAGLI P S L+L+NVG+N TRTGI++V+ MG ++ +++ + ++ V S++KGT
 Sbjct: 241 VAGLIAPNSRLVLQNVGINETRTGIIIDVIRAMGGKLEITEIDPVAKSATLIVESSDLKGT 300

40 Query: 304 IISGGLIPRLIDELPIIALLATQAQGTTCIKDAQELVKETDRIQVVDILNSMGANIK 363
 I G LIPRLIDELPIIALLATQAQ T IKDA+EL+VKETDRIQV D LNSMGA+I
 Sbjct: 301 EICGALIPRLIDELPIIALLATQAQGVTVIKDAEELVKETDRIQVVDALNSMGADITP 360

45 Query: 364 TADGMIKGPVLYGANTSTYGDHRIGMMTAIAALLVKQGQVHLDKEAIMTSYPTFFKD 423
 TADGMIK G + L+GA +T+GDHRIGMMTAIAALLV G+V LD+ EAI TSYP+FF D
 Sbjct: 361 TADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVADGEVELDRAEAINTSYPSFFDD 420

Query: 424 LERLCH 429
 LE L H
 Sbjct: 421 LESLIH 426

An alignment of the GAS and GBS proteins is shown below.

Identities = 269/424 (63%), Positives = 331/424 (77%)

50 Query: 5 MKLLTNANTLKGTRVPGDKSISHRAIIFGSISQGVTRIVDVLRGEDVLSTIEAFKQMGV 64
 MKL TNA L+GTI+VPGDKSISHRA+I G++++G TR+ +L+GEDVLSTI+AF+ +GV
 Sbjct: 4 MKLRTNAGPLQGTIQVPGDKSISHRAVILGAVAKGETRVKGLLKGEDVLSTIQAFRNLGV 63

55 Query: 65 LIEDGEIITIYKGKGFAGLTQPNLLDMGNSGTSMRLIAGVLAGQEFVMTVGDNSLSKR 124
 IE+ + + I G+GF GL P L+MGNSGTSMRLIAG+LAGQ F V M+GD SLSKR
 Sbjct: 64 RIEEKDDQLVIEGQGFQGLNAPCQTLNMGNSGTSMRLIAGLLAGQPFVSKMIGDESLSKR 123

60 Query: 125 PMDRIALPLSKMGARISGVTNRDLPLKLQGTIKLKPIFYHLPVASAQVKSALIFAALQT 184
 PMDRI PL +MG ISG T+R PPL+LQG + L+PI Y LP++SAQVKSAA+ AALQ
 Sbjct: 124 PMDRIVYPLKQMGVEISGETDRQFPPLQLQGNRNLPITYTLPISSAQVKSAILLAALQA 183

65 Query: 185 KGESLIVEKEQTRNHTEDMIRQFGGHLDIKDKIEIRLNGGQSLVGQDIRVPGDISSAAFVWI 244
 KG + +VEKE TRNHTE+MI+QFGG L + K I L G Q L Q+I VPGDISSAAFV+
 Sbjct: 184 KGTTQVVEKEITRNHTEEMIQQFGRLIVDGKRITLVGPQQLTAQEITVPGDISSAAFVWL 243

Query: 245 VAGLIIPNSHIILENVGINETRTGILDVVSVMGGKIKLSSVDNQVKSATLTVDYSHLQAT 304
 VAGLIIP S ++L+NVG+N TRTGIL+VV KMG +I ++ + + ++ V YS+++ T

-1424-

Sbjct: 244 VAGLIIPGSELLLNKVGVPNPTRTGILEVVEKMGAIQVYEDMNKKEQVTSIRVVYSNMKGT 303

Query: 305 HISGAMIPRLIDELPIALLATQAQGTTVIADAQELVKETDRIQVVESLKQMGADITA 364
ISG +IPRLIDELPIALLATQAQGT I DAQEL+VKETDRIQVV + L MGA+I A

5 Sbjct: 304 IISGGLIPRLIDELPIALLATQAQGTTCIKDAQELRVKETDRIQVVTDILNSMGANIK 363

Query: 365 TADGMIIRGNTPHHAASLDCHGDHRIGMMIAIAALLVKEGEVDLSGEEAINTSYPNFLEH 424
TADGMII+G T L+ A+ +GDHRIGMM AIAALLVK+G+V L EEAI TSYP F +

10 Sbjct: 364 TADGMIKGPVLYGANTSTYGDHRIGMMTAIAALLVKGQGVHLDKBEAINTSYPTFFKD 423

Query: 425 LEGL 428
LE L

Sbjct: 424 LERL 427

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1287

A DNA sequence (GBSx1364) was identified in *S.agalactiae* <SEQ ID 3959> which encodes the amino acid sequence <SEQ ID 3960>. Analysis of this protein sequence reveals the following:

20 Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -1.12 Transmembrane 6 - 22 (6 - 22)

25 ----- Final Results -----
bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAF20148 GB:AF208390 actinin-like protein [Entamoeba
histolytica]
Identities = 62/236 (26%), Positives = 107/236 (45%), Gaps = 38/236 (16%)

35 Query: 144 NYNSTNSSNPESMLFYEQKLTWLSTH----KNYYLDYK--VTPITYQNNELIPRKIELK- 196
N N + N + + L W+++ N+ D+K V + + +I+ +
Sbjct: 116 NANQQKNVNAKEEVVENNALLDWVNSFGLNVSNFSSDWKDGVALVKLTEAVSAGQIKFEQ 175

40 Query: 197 YVGIDKTGKLLPIFIGNKSTQDQFGI-----STVTLENTSPNATIDYLSGKAQN----- 245
+ G+D T ++ K +QF I + E P + + Y+S +
Sbjct: 176 FSGLDNTQMVIDC---QKLAYEQFKIPILMDVKDLVCERPDPKSIMTYVSVYKERYEQLL 232

45 Query: 246 TVLSAKEQRKLIKHEEEKRLAEK----KVEEEKAAAEQKKL-EEEQARLAAEAQ-RK 298
KE+++ IA+ E+E++ E+ + E+E+ A E Q++L EEQ RLA E Q RK
Sbjct: 233 VEKEQKEEQERTAREEQERKQKEEQERLAREEQERLAREEQERLAREEQERLAREEQERK 292

50 Query: 299 QKEEQARLAAAEQKKQETLVQEQTSQGYKRDYRGRWHPNGQYASKAEIAAAGLQW 354
QKEEQ RLA E Q++++ QE+ +Q +P Q + + AA W
Sbjct: 293 QKEEQERLAREEQERKQREEQERLNQ-----QQPTSQQLTFFSVQAADAW 338

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3961> which encodes the amino acid sequence <SEQ ID 3962>. Analysis of this protein sequence reveals the following:

Possible site: 41

55 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60

-1425-

The protein has homology with the following sequences in the databases:

```

>GP:CAA03161 GB:A49208 unnamed protein product [Streptococcus
pyogenes]
Identities = 54/222 (24%), Positives = 93/222 (41%), Gaps = 39/222 (17%)

5   Query: 44  HYKNTVSSKLLP--FTANYQLQLGELDNLNRA-----TFSHIQLQDRHETKDVRTKINYD 96
      +YK  +S++ P  F      +  +LD L R      T ++ ++ + +  K N +
Sbjct: 76  YYKTLGTSQITPALFPKAGDILYSKLDLGRTRTARGTLTYANVEGSYGVRQSFGK-NQN 134

10  Query: 97  PVGWHN-----YQFPYGDG-SKSSWVMNRGHLVGYQFCGLNDEPRNLVAMTAWLNTGAY 149
      P GW      Y+  + +G S      NR HL+      G      + + + A  T
Sbjct: 135  PAGWTGNPNHVKYKIEWLNGLSYVGDFWNRSHLIADSLGG-----DALRVNAVGTGTRTQ 188

15  Query: 150 SGANDSNPEGMLYYENRLDSWLALHPDFWLDYKVTPIYSGNEVVPRQIELQYVGIDSSGE 209
      +      GM Y E R  WL  + D +L Y+V PIY+ +E++PR +
Sbjct: 189  NVGGRDQKGGMRYTEQRAQEWLEANRDGYLYYEVAPIYNADELIPRAV----- 236

Query: 210  LLTIRLNSNKESIDENGVTITVILENSAPNINLDYLNGTATPK 251
      + + S+  +I+E      V++ N+A      ++Y NGT T K
20  Sbjct: 237  --VSMQSSDNTINEK----VLVYNTANGYTTINYHNGTPTQK 272

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/245 (47%), Positives = 166/245 (67%), Gaps = 4/245 (1%)

```

25  Query: 2    KRKQFIKLG IATLLTVISLYTPINLATNHTTENIVTAQ EY--KTKENGTL PFKHKRQLVL 59
      K+K  +  +  LL++      ++ A  T  N+  A  +  T  +  LPF  QL L
Sbjct: 5    KQKASLLTAVLLLLLSLSITTTITVDAARVRTYPNVSHANTHYKNTVSSKLLPFTANYQLQL 64

30  Query: 60  GELDDKGRATFAHIQLKVKDEPKKKRVKRLKTPVGVHNFKFYNDGTQKAWLMSRGRLI 119
      GELD+  RATF+HIQL+  + E K  R K  +  PVGWHN++F Y DG++ +W+M+RG L+
Sbjct: 65  GELDNLNRATFSHIQLQDRHETKDVRTK- INYDPVGVWHNYQFPYGDGSKSSWVMNRGHLV 123

Query: 120  CHQFSGLNNERKNLVPMTNWLNTGNYNSTNSSNPESMLFYEKQLKTWLSTHKNYLDYKV 179
      +QF GLN+E +NLV MT WLNTG Y+  N SNPE ML+YE +L +WL+ H +++LDYKV
35  Sbjct: 124  GYQFCGLNDEPRNLVAMTAWLNTGAYSGANDSNPEGMLYYENRLDSWLALHPDFWLDYKV 183

Query: 180  TPIYQNNELIPRKIELKYVGIDKTGKLLPIFI-GNKSTQDQFGISTVTLENTSPNATIDY 238
      TPIY  NE++PR+IEL+YVGID +G+LL I +  NK + D+ G++TV LEN++PN +DY
40  Sbjct: 184  TPIYSGNEVVPRQIELQYVGIDSSGELLTIRLNSNKESIDENGVTITVILENSAPNINLDY 243

Query: 239  LSGKA 243
      L+G A
Sbjct: 244  LNGTA 248

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7263> which encodes amino acid sequence <SEQ ID 7264>. An alignment of the GAS and GBS sequences follows:

Score = 58.9 bits (140), Expect = 2e-11
Identities = 34/103 (33%), Positives = 55/103 (53%), Gaps = 1/103 (0%)

```

50  Query: 1    MPFKTNLKG ILLYAMFMASIFLLVLQVYLSQVTALHKEYQAQTDYVKARLIAEIVYQD- 59
      M  K  LKAGILL A+ +A++F LVLQ YL+++ A  ++Y +Q +  KA L A++ Y+
Sbjct: 1    MILKKKLKAGILLQAIVLAAVFTLVLFQYLARILATERQYHSQIEASKAYLTAQLAYKTI 60

Query: 60  HRYKASNPVFFKGGQVICRERKERWMLIVKLDQQRQYQFEYLK 102
      S  +F GG      +      + V LD+  Y  ++ +
55  Sbjct: 61  EGDSISGKCYFTGGYASYLQEGNYLQVKVTLDKGGNYNHKFYR 103

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1426-

Example 1288

A DNA sequence (GBSx1365) was identified in *S.agalactiae* <SEQ ID 3963> which encodes the amino acid sequence <SEQ ID 3964>. This protein is predicted to be enolase (eno). Analysis of this protein sequence reveals the following:

```

5   Possible site: 43
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3025(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15 >GP:BAA81815 GB:AB029313 enolase [Streptococcus intermedius]
   Identities = 396/435 (91%), Positives = 414/435 (95%), Gaps = 1/435 (0%)

   Query: 1   MSIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG 60
           MSIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG
   Sbjct: 1   MSIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG 60

20   Query: 61  GLGTQKAVDNVNNVIAEAIIGYDVRDQQAIDRAMIALDGTGNKGLGANAILGVSIIVAR 120
           GLGTQKAVDNVNN+IAEA+IGYDVRDQQAIDRAMIALDGTGNKGLGANAILGVSIIVAR
   Sbjct: 61  GLGTQKAVDNVNNIIAEAVIGYDVRDQQAIDRAMIALDGTGNKGLGANAILGVSIIVAR 120

25   Query: 121 AAADYLEVPLYSYLGGFNTKVLPTPMNIIINGGSHSDAPIAQEFMIMFVGAPTFKEALR 180
           AAADYLE+PLYSYLGGFNTKVLPTPMNIIINGGSHSDAPIAQEFMI+P GAPTFKEALR
   Sbjct: 121 AAADYLEIPLYSYLGGFNTKVLPTPMNIIINGGSHSDAPIAQEFMIVPAGAPTFKEALR 180

30   Query: 181 WGAEVFHALKKILKRGLETAVGDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI 240
           WGAE+FHALKKILK RGL TAVGDEGGFAP+F+GTEDGVETIL AIEAAGY G++ + +
   Sbjct: 181 WGAEIFHALKKILKSRGLATAVGDEGGFAPRFDGTEDGVETILAAIEAAGYVPGKD-VFL 239

   Query: 241 GFDCASSEFYDAERKVVYDYSKFEGEGGAVRTAAEQIDYLEELVNKYPIITIEDGMDENDW 300
           GFDCASSEFYD ERKVYDY+KFEGEG AVRTA EQIDYLEELVNKYPIITIEDGMDENDW
35   Sbjct: 240 GFDCASSEFYDKERKVYDYTKFEGEAAVRTADEQIDYLEELVNKYPIITIEDGMDENDW 299

   Query: 301 DGWKALTERLGGRVQLVGDDFFVTNTDYLARGIKEEAANSILIKVNQIGTLTETFEAIEIEM 360
           DGWK LTERLG +VQ VGDDFFVTNT YL +GI E ANSILIKVNQIGTLTETF+AIEIEM
   Sbjct: 300 DGWKKLTERLGKKVQPVGDDFFVTNTSYLEKGINEACANSILIKVNQIGTLTETFDIEIEM 359

40   Query: 361 AKEAGYTAVVSHRSGETEDSTIADIATAVATNAGQIKTGSLSRTDRIAKYNQLLRIEDQLGE 420
           AKEAGYTAVVSHRSGETEDSTIADIATAV NAGQIKTGSLSRTDRIAKYNQLLRIEDQLGE
   Sbjct: 360 AKEAGYTAVVSHRSGETEDSTIADIATAVATNAGQIKTGSLSRTDRIAKYNQLLRIEDQLGE 419

45   Query: 421 VAQYKGIKSFYNLKK 435
           VA+Y+G+KSFYNL K
   Sbjct: 420 VAEYRGLKSFYNLSK 434

```

Proteins in the glycolysis/gluconeogenesis pathway have been experimentally detected on the surface of *Streptococci*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3965> which encodes the amino acid sequence <SEQ ID 3966>. Analysis of this protein sequence reveals the following:

```

55   Possible site: 43
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3025(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
60      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```


-1427-

The protein has homology with the following sequences in the databases:

>GP:BAA81815 GB:AB029313 enolase [Streptococcus intermedius]
Identities = 396/435 (91%), Positives = 415/435 (95%), Gaps = 1/435 (0%)

5 Query: 1 MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYL 60
MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRY
Sbjct: 1 MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG 60

10 Query: 61 GLGTQKAVDNVNNIIAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR 120
GLGTQKAVDNVNNIIAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR
Sbjct: 61 GLGTQKAVDNVNNIIAEAVIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR 120

15 Query: 121 AAADYLEVPLYTYLGGFNTKVLPTPMMNIINGGSHSDAPIAQEFMIMPVGAPT FKEGLR 180
AAADYLE+PLY+YLGGFNTKVLPTPMMNIINGGSHSDAPIAQEFMI+P GAPTFKE LR
Sbjct: 121 AAADYLEIPLYSYLGGFNTKVLPTPMMNIINGGSHSDAPIAQEFMIVPAGAPT FKEALR 180

20 Query: 181 WGAEVFHALKKILKERGLVTAVGDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI 240
WGAE+FHALKKILK RGL TAVGDEGGFAP+F+GTEDGVETIL AIEAAGY G++ ++
Sbjct: 181 WGAEIFHALKKILKSRGLATAVGDEGGFAPRFDGTEDGVETILAAIEAAGYVPGKD-VFL 239

25 Query: 241 GFDCASSEFYDKERKVYDYTKFELEGAAVRTSAEQVDYLEELVNKYPIITIEDGMDENDW 300
GFDCASSEFYDKERKVYDYTKFELEGAAVRT+ EQ+DYLEELVNKYPIITIEDGMDENDW
Sbjct: 240 GFDCASSEFYDKERKVYDYTKFELEGAAVRTADEQIDYLEELVNKYPIITIEDGMDENDW 299

30 Query: 301 DGWKVLTERLGKRVQLVGDDFFVTINT EY LARGIKENAANSILIKVNQIGTLTETFEA IEM 360
DGWK LTERLGK+VQ VGDDFFVTINT YL +GI E ANSILIKVNQIGTLTETF+A IEM
Sbjct: 300 DGWKKLTERLGKRVQPVGDDFFVTINTSYLEKGINEACANSILIKVNQIGTLTETFD A IEM 359

35 Query: 361 AKEAGYTAVVSHRSGETEDSTIADIAVATNAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE 420
AKEAGYTAVVSHRSGETEDSTIADIAVA NAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE
Sbjct: 360 AKEAGYTAVVSHRSGETEDSTIADIAVAANAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE 419

40 Query: 421 VAQYKGIKSFYNLKK 435
VA+Y+G+KSFYNL K
Sbjct: 420 VAEYRGLKSFYNLSK 434

An alignment of the GAS and GBS proteins is shown below.

Identities = 421/435 (96%), Positives = 427/435 (97%)

40 Query: 1 MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG 60
MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRY
Sbjct: 1 MSIIITDVYAREVLDSRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYL 60

45 Query: 61 GLGTQKAVDNVNNVIAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR 120
GLGTQKAVDNVNN+IAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR
Sbjct: 61 GLGTQKAVDNVNNIIAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR 120

50 Query: 121 AAADYLEVPLYSYLGGFNTKVLPTPMMNIINGGSHSDAPIAQEFMIMPVGAPT FKEALR 180
AAADYLEVPLY+YLGGFNTKVLPTPMMNIINGGSHSDAPIAQEFMIMPVGAPT FKE LR
Sbjct: 121 AAADYLEVPLYTYLGGFNTKVLPTPMMNIINGGSHSDAPIAQEFMIMPVGAPT FKEGLR 180

55 Query: 181 WGAEVFHALKKILKERGLETA V GDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI 240
WGAEVFHALKKILKERGL TAVGDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI
Sbjct: 181 WGAEVFHALKKILKERGLVTAVGDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI 240

60 Query: 241 GFDCASSEFYDAERKVYDYSKFEGEGGAVRTAAEQIDYLEELVNKYPIITIEDGMDENDW 300
GFDCASSEFYD ERKVYDY+KFEGEG AVRT+AEQ+DYLEELVNKYPIITIEDGMDENDW
Sbjct: 241 GFDCASSEFYDKERKVYDYTKFELEGAAVRTSAEQVDYLEELVNKYPIITIEDGMDENDW 300

65 Query: 301 DGWKALTERLGGRVQLVGDDFFVTINTDY LARGIKEEAANSILIKVNQIGTLTETFEA IEM 360
DGWK LTERLG RVQLVGDDFFVTINT+YLARGIKE AANSILIKVNQIGTLTETFEA IEM
Sbjct: 301 DGWKVLTERLGKRVQLVGDDFFVTINT EY LARGIKENAANSILIKVNQIGTLTETFEA IEM 360

Query: 361 AKEAGYTAVVSHRSGETEDSTIADIAVATNAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE 420
AKEAGYTAVVSHRSGETEDSTIADIAVATNAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE

-1428-

Sbjct: 361 AKEAGYTAVVSHRSGETEDSTLIADIAVATNAGQIKTGSLSRTDRIAKYNQLLRIEDQLGE 420

Query: 421 VAQYKGIKSFYNLKK 435

VAQYKGIKSFYNLKK

5 Sbjct: 421 VAQYKGIKSFYNLKK 435

SEQ ID 3964 (GBS311) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 3; MW 51kDa).

GBS311-His was purified as shown in Figure 203, lane 10.

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1289

A DNA sequence (GBSx1366) was identified in *S.agalactiae* <SEQ ID 3967> which encodes the amino acid sequence <SEQ ID 3968>. Analysis of this protein sequence reveals the following:

15 Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.1998(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1290

A DNA sequence (GBSx1367) was identified in *S.agalactiae* <SEQ ID 3969> which encodes the amino acid sequence <SEQ ID 3970>. This protein is predicted to be di-/tripeptide transporter. Analysis of this protein sequence reveals the following:

30 Possible site: 54
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -14.33	Transmembrane	93 - 109 (87 - 122)
INTEGRAL	Likelihood = -9.02	Transmembrane	117 - 133 (110 - 141)
35 INTEGRAL	Likelihood = -8.44	Transmembrane	333 - 349 (328 - 353)
INTEGRAL	Likelihood = -5.84	Transmembrane	19 - 35 (17 - 38)
INTEGRAL	Likelihood = -3.08	Transmembrane	151 - 167 (151 - 167)
INTEGRAL	Likelihood = -2.55	Transmembrane	264 - 280 (264 - 281)
INTEGRAL	Likelihood = -2.28	Transmembrane	44 - 60 (44 - 60)
40 INTEGRAL	Likelihood = -2.02	Transmembrane	238 - 254 (238 - 255)

----- Final Results -----

45 bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9395> which encodes amino acid sequence <SEQ ID 9396> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1429-

>GP:CAB12175 GB:Z99106 similar to di-tripeptide ABC transporter
(membrane protein) [Bacillus subtilis]

Identities = 175/359 (48%), Positives = 254/359 (70%), Gaps = 9/359 (2%)

5 Query: 1 MVGNLYGENDSRRDAGFSIFVFGINLGAFISPIVVGYLQGEVNFHLGFSLAAIGMFFGLL 60
+VG+LY + D RRD+GFSIF GINLG ++P++VG LGQ+ N+HLGF AA+GM GL+
Sbjct: 142 VVGDLTYTKEDPRRDSGFSIFYMGINLGGLLAPLIVGTLGQKYNHYLGFCAAAGVGMMLGLI 201

10 Query: 61 QYTLDGKKYLTEESLRPNPDPLSPEEKSSLYKKVGLILIGIVIVLILLHLMHMLTIEVIID 120
+ L KK L +PLS +KS++ +G+I++ I +++ + +LTI+ ID
Sbjct: 202 VFPLTRKKNLGLAGSNVPNPLS--KKSAGTGIGVIIAIAVIISVQ--TGVLTIKRFID 257

15 Query: 121 IFSIIAIAIPIIYFIKILSSKKISSVERSrvWAYIPLFIASILFWSIEEQGSVVLALFAD 180
+ SI+ I IP+IYFI + +SKK E+SR+ AY+PLFI +++FW+I+EQG+ +LA++AD
Sbjct: 258 LVSILGILIPVIYFIIMFTSKKADKTEKSRLAAYVPLFIGAVMFVAIQEQGATILAVYAD 317

20 Query: 181 EQTKLYLNFFGGHHINFPSYFQSMNPLFIMLYVPFFAWLWAKWGSQPSPPKFAYGLFF 240
E+ +L L F SS+FQS+NPLF++++ P FAWLW K G +QPS+P KF+ G+
Sbjct: 318 ERIRLSLGGF---ELQSSWFQSLNPLFVVIFAPIFAWLMMKLGKRPSTPVKFSIGIIL 373

25 Query: 241 AGASFLWMMLPGLLFGVNAKVSPWLWTMSWAIVIVGEMLISPVGLSATS KLAPKAFOAQM 300
AG SF+ M+ P + G A VSPLWL +S+ +V++GE+ +SPVGLS T+KLAP AF AQ
Sbjct: 374 AGLSFIIMVFPAMQ-GKEALVSPLWLVLVSFLLVVLGELCLSPVGLSVTTKLAPAAFSQQT 432

30 Query: 301 MSIWFLSNAAAQAINAQIVKLYTPDTQTLYYGVVGGITVVFGEFILLFYVPRIEKLMSGV 359
MS+WFL+NAAAQAINAQ+ L+ +T+Y+G +G I++V G ILL P I++ M GV
Sbjct: 433 MSWFLTNAAAQAINAQVAGLFDKIPETMYFGTIGLISIVLGGILLLLSPVIKRAMKGV 491

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1291

A DNA sequence (GBSx1369) was identified in *S.agalactiae* <SEQ ID 3971> which encodes the amino acid sequence <SEQ ID 3972>. Analysis of this protein sequence reveals the following:

35 Possible site: 37
>>> Seems to have an uncleavable N-term signal seq

40 ----- Final Results -----
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1292

A DNA sequence (GBSx1370) was identified in *S.agalactiae* <SEQ ID 3973> which encodes the amino acid sequence <SEQ ID 3974>. Analysis of this protein sequence reveals the following:

50 Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2485(Affirmative) < succ>

-1430-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAF61315 GB:U96166 unknown [Streptococcus cristatus]
 Identities = 181/442 (40%), Positives = 270/442 (60%), Gaps = 2/442 (0%)

Query: 1 MINLFDSYTQSSWDLHFSLIKSGVINPTIALNDDGFLPDDVTSPYLYYTGFAGTGRPL 60
 MI LFD Y Q+S+DL SL +G P + + DDG+L DV SPY Y+TG T GRP+
 10 Sbjct: 1 MICLFDRYDQASFDLLRLSKATGLDCPVVVVQDDGYLSPDVESFYFTGDLDTPEGRPI 60

Query: 61 YYNELRVPTWEIIGFSSGADIVDLGVKKGRITYANPNHKRLIKEVDWFDEQGRVILKDR 120
 Y+N + P WEI + +I+D+G K+ I Y P H+R ++ V+W D +G+V D
 15 Sbjct: 61 YFNLVPKPHLWEIRSSNVNGEILDMGKKRANIFYRQPTHERRVRAVEWLDTEGQVRAADI 120

Query: 121 FNKFGFCFAQTFFYNADGQAIQTSYYNKDRQEVISENHMTGDIILNDNNQKVFVSKVEFV 180
 +N+ G FAQ Y+ + T Y+++ VI ENH+TGD IL + +FKSK Efv
 20 Sbjct: 121 YNRGRLEFAQITYDQTQRPHTRYFDQSNVVMENHLTGDIILTLEGKRHIFKSKQEFV 180

Query: 181 INYLQEAQFNLDRIFFYNLSLSTPFLVSFYLL--NRLESKDVLFWQEPLVDDIPGNMRLLLNN 238
 + YLQ ++ DRI YNSL+TPFLV++ L ++DVLFWQEP+ + +PGNM++ +
 25 Sbjct: 181 VFYLQYRGYDTRIIYNSLATPFLVAYALRPKNGRAEDVLFWQEPIGEALPGNMKVAMKM 240

Query: 239 PSPNTKIVIQSYEAYANAMRLLTDEEQKQVSFLGFMYPLKETEKLNQALILITNSDQIEA 298
 P N +I +Q + Y L T EE+ +G++Y + ++ +ALILITNSDQ+E
 30 Sbjct: 241 PHRNIRIAVQDRQVYEKIQSLATPEEKVYFHNIGYIYDYQRLNNMNPEALILITNSDQLEQ 300

Query: 299 LESLVTSLPNLTFNIGALTEMSSDLNMFNGKYDNVVLYPNITTNIQYLSNICAFYLDINH 358
 +E L+T LPN+ F+IGA+TEMS LM +Y NV LYPNI ++ L C YLDIN
 35 Sbjct: 301 IEQLLTQLPNVHFHIGAITEMSGHLMGLNRYPNVSLYPNIRPAKVAELFERCDLYLDINI 360

Query: 359 HNEILSAVRSFAFEHQQLIFAFEETSHQIRFVSPKNIFFPKKDIFTFISHLOPLIGNKCNIE 418
 +EIL+A R+AFE+ LI +F T H RF++ +I+ +++ + +Q + + +E
 40 Sbjct: 361 SDEILNACRTAFENMLILSFTINTCHSRRFIADDDHIYAPENVSGMVDKIQSALAHSSEME 420

Query: 419 KALKQOLEDCHVSSSTQYQSVI 440
 AL +Q + + +S QY+++I
 45 Sbjct: 421 AALTRQKQAANQASLEQYKAI 442

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1293

45 A DNA sequence (GBSx1371) was identified in *S.agalactiae* <SEQ ID 3975> which encodes the amino acid sequence <SEQ ID 3976>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.06 Transmembrane 405 - 421 (404 - 422)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA94320 GB:AB033763 hypothetical protein [Staphylococcus aureus]
 Identities = 66/195 (33%), Positives = 99/195 (49%), Gaps = 9/195 (4%)

-1431-

Query: 259 NYDYQFTNANRFDFFITSTDKQTELEQQFKQFTNHNPRITIPVGSID----NLKMPM 314
 N Y + F N NR+ I ST +Q + N+ + TIPVG ID NLK
 Sbjct: 15 NTKYKHVFNLNRYSGIIVSTKQQ----QLDISARINNEIPVHTIPVGYIDHFTNLKRNN 70

5 Query: 315 DNRRPYSILTASRLASEKHVDLVRVIRIREILPEVTFDIYSGSGGEEKIRNIINAANA 374
 + I++ +R + EK ++ + V ++ + P + +YG G EEEK + +I N
 Sbjct: 71 HSINNKKIISVARYSPEKQLNHQIELVSKLIKEFPNIRLHLVYFGKEEEKYKQLITEYNL 130

10 Query: 375 TEYIRLMG-HKNLSNVYQNYELYLTASKSEGFGLTLLAIGAGLPLIGFDVRYGNQTFIK 433
 + L G +NLS Q+ + L S EGF L LLE I G+P +G++ +YG I
 Sbjct: 131 ENNVFLRGFRRLNSAEIQDAYMSLITSNMEGFNLGLETTITTEGIPPVGYNKYGPSELIL 190

Query: 434 DGENGYLIPRFDMD 448
 + ENGYLI + D D+
 15 Sbjct: 191 NNENGYLINKNDKDE 205

SEQ ID 3976 (GBS426) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 4; MW 58.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 3; MW 84kDa).

20 GBS426-GST was purified as shown in Figure 220, lane 5.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1294

A DNA sequence (GBSx1372) was identified in *S.agalactiae* <SEQ ID 3977> which encodes the amino acid sequence <SEQ ID 3978>. This protein is predicted to be preprotein translocase *seca* subunit (*secA*). Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.69 Transmembrane 75 - 91 (75 - 91)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44957 GB:U56901 involved in protein export [Bacillus subtilis]
 Identities = 336/794 (42%), Positives = 506/794 (63%), Gaps = 29/794 (3%)

40 Query: 5 NSLFSLDKKRLKKLQRTLTNTINSKGMATLSNEELQAKTTEFRKRLVNGETLDDICAE 64
 N +F K+ L + ++ N I++++G LS++ L+ KT EF++RL G T DD+ EA
 Sbjct: 6 NKMFDPTKRTLNRYEKIANDIDAIRGDYENLSDDALKHKTIEFKERLEKGATTDDLIVEA 65

45 Query: 65 FAVVREADERVGLFPYDVQVIGGLVLHQNTAEMKTGEGKTLTATMPLYLNALEGKGM 124
 FAVVREA RV G+FP+ VQ++GG+ LH GN AEMKTGEGKTLT+T+P+YLNAL GK
 Sbjct: 66 FAVVREASRRVTGMFPFKVQLMGVALHDGNTAEMKTGEGKTLTSTLPVYLNALTGKGVH 125

Query: 125 LLTMSYLAIRDAEMGKVYRFLGSLVGVGSDNEEEDRDAATKRAYSSDIVYSTSSAL 184
 ++T N YLA RDAE+MGK++ FLGL+VG+ ++ +++ KR Y++DI YST++ L
 50 Sbjct: 126 VVTVNEYLASRDAEQMGKIFEFGLT+VGLNLNSMSKDE-----KREAYAADITYSTNNEL 180

Query: 185 GFDYLIDNLASSKSQKYMPLHYAIVDEADAVLLDMAQTPLVISGSPRVQSNLYKIADEL 244
 GFDYL DN+ K Q LH+A++DE D++L+D A+TPL+ISG + LY A+
 Sbjct: 181 GFDYL RDNMVLKYEQMVQRPLHFAVIDEVDLSILIDEARTPLIISGQAAKSTKLYVQANAF 240

55 Query: 245 ILSFEEQVDYDFDKERQEVWIKNQGVREAERYFRI PHFYKQSNRELVRHLNLSLKAHKLF 304
 + + + + DY +D + + V + +G+ +AE+ F I + + + L H+N +LKAH
 Sbjct: 241 VRTLKAEKDYTYDIKTAVQLTEEGMTKAEKAFGIDNLFVDVKHVALNHHINQALKAHVAM 300

-1432-

5 Query: 305 ERGKDYVVDGEIKLLDATNGRVLEGTKLQGGVHQAEQKEHLNVTPESTRAMASITYQNL 364
 ++ DYVV+DG++ ++D+ GR+++G + G+HQAIE KE L + ES +A+IT+QN
 Sbjct: 301 QKDVDYVVEDGQVVIVDSFTGRLMKGRRYSEGLHQAIEAKEGLEIQNESMTLATITTFQNY 360

10 Query: 365 FRMFTKLAGMTGTGKTAEKEFIEVDMEVVRIPITNSPVRRIDYDPDKIYTTLPKIHATIE 424
 FRM+ KLAGMTGT KT E+EF +Y+M+VV IPTN PV R D PD IY T+ K A E
 Sbjct: 361 FRMYEKLAGMTGTAKTEEEEFNRNIYNMQVVTIPTNRPVVRDDRPDLIYRTMEGKFKFAVAE 420

15 Query: 425 FVKQVHDTGQPILLVAGSVRMSSELFSELLLSGIPHSLNNAQSAVKEAQMIAEAGQKGAV 484
 V Q + TGQP+L+ +V SEL S+LL GIPH +LNA++ +EAQ+I EAGQKGAV
 Sbjct: 421 DVAQRYMTGQPVLVGTAVAVETSELISKLLKNKGIPHQVLNAKNHEREAQIIEEAGQKGAV 480

20 Query: 485 TVATNMAGRGTDIKLGKGVSELGLAVIGTERMSQRMQLRGRSGRQGDIGFSQFFVS 544
 T+ATNMAGRGTDIKLG+GV ELGGLAV+GTER +S+R+D QLRGRSGRQGD G +QF++S
 Sbjct: 481 TIATNMAGRGTDIKLGEGVKELGGLAVVGTERHESRRIDNQLRGRSGRQGDGPITQFYLS 540

25 Query: 545 FEDDLMIESGPKWAQDYFRKNRDKVNPEKPKALGQRRFQKLFQQTQOEASDGKGESARSQT 604
 ED+LM G + D+ + + + + +Q+ +G +R Q
 Sbjct: 541 MEDELMRRFGAERTMAML---DRFGMDDSTPIQSKMVSRAVESSQKRVEGNNFDSRKQL 596

30 Query: 605 IEFDSSVQLQREYVYRERNALINGESGHFSRQIIDTVISSFI-----AYLDGEVEKEEL 659
 +++D ++ QRE +Y++R +I+ E + R+I++ +I S + AY E EE
 Sbjct: 597 LQYDDVLRQQREVIYKQRFVIDSE---NLREIVENMIKSSLERATAAYTPREELPEE- 651

35 Query: 660 IFEVNRFI-FDNMSYNLQGISKEMSL--EEIKNYLFKIADEILREKHNLLGDSFG----- 711
 ++++ + N +Y +G ++ + +E L I D I+ K+N + FG
 Sbjct: 652 -WKLDGLVDLINTTYLDEGALEKSDIFGKEPDEMLELIMDRII-TKYNEKEEQFGKEQMR 709

40 Query: 712 DFERTAALKAIDEAWIEEVDYLQQLRTVATARQTAQRNPVFYHKEAYKSYNIMKKEIRE 771
 +FE+ L+A+D W++ +D + QLR R AQ NP+ EY E + + M + I +
 Sbjct: 710 EFEKVVILRAVDSKMDHIDAMDQLRQGIHLRAYAQTNPRLREYQMEGFAMFEHMIESIED 769

Query: 772 QTFRNLLLSEVSFN 785
 + + ++ +E+ N
 Sbjct: 770 EVAKFVMKAEIENN 783

There is also homology to SEQ ID 3620.

40 SEQ ID 3978 (GBS425) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 3; MW 91kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 2; MW 116kDa).

GBS425-GST was purified as shown in Figure 220, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

45 Example 1295

A DNA sequence (GBSx1373) was identified in *S.agalactiae* <SEQ ID 3979> which encodes the amino acid sequence <SEQ ID 3980>. Analysis of this protein sequence reveals the following:

50 Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3827(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

-1433-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1296

- 5 A DNA sequence (GBSx1374) was identified in *S.agalactiae* <SEQ ID 3981> which encodes the amino acid sequence <SEQ ID 3982>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2683(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 A related GBS nucleic acid sequence <SEQ ID 10001> which encodes amino acid sequence <SEQ ID 10002> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1297

A DNA sequence (GBSx1375) was identified in *S.agalactiae* <SEQ ID 3983> which encodes the amino acid sequence <SEQ ID 3984>. Analysis of this protein sequence reveals the following:

25 Possible site: 31
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.5410(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1298

A DNA sequence (GBSx1376) was identified in *S.agalactiae* <SEQ ID 3985> which encodes the amino acid sequence <SEQ ID 3986>. This protein is predicted to be preprotein translocase secy subunit. Analysis of this protein sequence reveals the following:

40 Possible site: 59
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -9.92 Transmembrane 287 - 303 (278 - 309)
INTEGRAL Likelihood = -9.08 Transmembrane 191 - 207 (186 - 210)
INTEGRAL Likelihood = -8.44 Transmembrane 104 - 120 (101 - 123)
45 INTEGRAL Likelihood = -8.23 Transmembrane 11 - 27 (9 - 41)

-1434-

INTEGRAL Likelihood = -3.93 Transmembrane 133 - 149 (129 - 150)
 INTEGRAL Likelihood = -3.19 Transmembrane 347 - 363 (344 - 364)
 INTEGRAL Likelihood = -2.97 Transmembrane 158 - 174 (155 - 174)
 INTEGRAL Likelihood = -1.54 Transmembrane 246 - 262 (245 - 262)
 INTEGRAL Likelihood = -0.90 Transmembrane 372 - 388 (372 - 388)
 INTEGRAL Likelihood = -0.85 Transmembrane 64 - 80 (64 - 81)

----- Final Results -----

bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF30659 GB:AE002122 preprotein translocase [Ureaplasma urealyticum]
 Identities = 105/422 (24%), Positives = 213/422 (49%), Gaps = 49/422 (11%)

Query: 2 KLLYIFEKNIILRKILITFSLIIIFLLGRYVPIPGVLISAYKGQDNNFATLYSTVTGGNL 61
 +LL IF+ +L +++T S++I+F +G +P+P + ++ G +F ++ + + GG L
 Sbjct: 13 QLLMIFKNKKVLVALIVTSLILILFRIGSVIPMPYIKLNGNFGNQGSFFSIINLLGGGGL 72

Query: 62 SQVGVFSLGIGPMTTMMILLRLFT-----IGKYSSGVSQKVVQFRQNVVMLVIAII 112
 SQ +F++GIGP +T I+++L + + K +K++ + ++ L +A++
 Sbjct: 73 SQFSLFAIGIGPYITAQIIMQLLSSELVPLAKLSKSGERGRKKIEVITR-IITLPLAVM 131

Query: 113 QGLAITISFYHNGFSL-----TKLLLATMI--LVTGAYIISWIGNLNAEYFGF- 159
 Q + I NGF + L T I +V G YI ++ +L ++ G G
 Sbjct: 132 QAVIIINLMTRANGFISIVSNAPFAIGSPLFVVTYIFLMVGGTYISLFLADLISKKGVG 191

Query: 160 GMTILVVVGMVLVGQFNNIPLIFELF-----QDGYQLAILFLLWTLVAMYLMITFERSE 213
 G+T+L++ G++ FN+ IF + + IL++L+ ++ + ++ S
 Sbjct: 192 GITLLILITGIVASLFNFHFAIFSNLGLSLTSSKVSQIIGFILYILFYIMILIGVVFVNNS 251

Query: 214 YRIPVMRTS-----IHNRLVDDAYMPIKVNASGGMAFMYVYTLIMFPQYIIILLRSIFPT 268
 +IPV +T H +L. ++PIK+ +G M ++ ++L P + L
 Sbjct: 252 RKIPVQQTGQALILDHEKL---PFLPIKIMTAGVMPVIFASSVLAIPAQVAEFLDK---Q 305

Query: 269 NPDITSYNDYFSLSSIQGVVIYIMLMLVLSVAFTFVNIDPTKISEAMRESGDFIPNYRPG 328
 + ++YF + S G+ IY++L+L+ + F++V ++P K++E +++G FIP + G
 Sbjct: 306 SMGYVVIHNYFIVDSWTGLAIYVVLILLFTFFFSYVQLNPPKMAEDIKKAGRFPVGVQVG 365

Query: 329 KETQSYLSKICYLFGTFSGFFMAFLGGVPLLFLALGNDDLr-----TVSSMTGIFMM 379
 +T+ +++K+ Y +AFL +P L AL + T+ T I +M
 Sbjct: 366 MDTEKHITKVIYRVNWIGAPILAFPLACPLHLVALVAKTINHGIPVIQIPSTIFGGTSIIM 425

Query: 380 IT 381
 +T
 Sbjct: 426 VT 427

There is also homology to SEQ ID 3988.

A related GBS gene <SEQ ID 8783> and protein <SEQ ID 8784> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 6.32
 GvH: Signal Score (-7.5): -4.07
 Possible site: 59

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 10 value: -9.92 threshold: 0.0

INTEGRAL Likelihood = -9.92 Transmembrane 287 - 303 (278 - 309)
 INTEGRAL Likelihood = -9.08 Transmembrane 191 - 207 (186 - 210)
 INTEGRAL Likelihood = -8.44 Transmembrane 104 - 120 (101 - 123)
 INTEGRAL Likelihood = -8.23 Transmembrane 11 - 27 (9 - 41)
 INTEGRAL Likelihood = -3.93 Transmembrane 133 - 149 (129 - 150)
 INTEGRAL Likelihood = -3.19 Transmembrane 347 - 363 (344 - 364)
 INTEGRAL Likelihood = -2.97 Transmembrane 158 - 174 (155 - 174)

INTEGRAL	Likelihood = -1.54	Transmembrane	246 - 262 (245 - 262)
INTEGRAL	Likelihood = -0.90	Transmembrane	372 - 388 (372 - 388)
INTEGRAL	Likelihood = -0.85	Transmembrane	64 - 80 (64 - 81)
PERIPHERAL	Likelihood = 8.65	28	
modified ALOM score: 2.48			

----- Final Results -----

```

bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

```

ORF02350(316 - 1500 of 1827)
EGAD|6621|6420(8 - 426 of 431) preprotein translocase secY subunit {Bacillus sp.}
SP|P38375|SECY_BACHD PREPROTEIN TRANSLOCASE SECY SUBUNIT. GP|484251|dbj|BAA01191.1||D10360
secretion protein Y {Bacillus sp.} PIR|B44859|B44859 preprotein translocase secY - Bacillus
sp.
%Match = 12.1
%Identity = 26.8 %Similarity = 55.4
Matches = 109 Mismatches = 165 Conservative Sub.s = 116

```

57 87 117 147 177 207 237 267
EVNIVVDRCITEGKTIYGIRRAKDKNOYISFEETMDDFEYLCDTIKONR*SRRVMT*ILKSIFLILKLT*TKLTT*SVLS*

297 327 357 387 441 471 501
REQIDREREIPLKLLYIFEKNIILRKILITFSLIIIFLLGRYPVPIGV--LISAYKGQDNNFATLYSTVTGGNLSQGVGF
 || : ||::: | : ::: | : | ::: || : : | | | | : | | | : |
MFTISINIFRVGDLRRKVIPTLLMLLIVFRIGSGFIPVPGTNRREVLDQANAFGFL-NTFGGGGALGNFSIF
 10 20 30 40 50 60 70

[illegible]

729 759 786 816 846 870 888 918
LATMILVTGAYIIISWIGNLNABYGF-GMTILVVVGMLVQFNFIPLIFEL-FQD-GYQL----AIIFLFWTLVAMYL
: :| : :| : :| | ||::: :| : :|| : || | || : ||| : ||| : ||| : ||| :
FIALVLTAGTAFLMWLGEQITAKGVNGISIIIFAGIAAGIPNGLNIYSTRIQDAGEFLFNIVVILLALAILAIIVG
160 170 180 190 200 210 220 230

966 1023 1053 1083 1113 1143
ITFERSEYR-IPVM---RTSIHNLRLVDDA-YMPIKVNASGGMAFMVYVYTLMLFPQYIIILLRSIFPTNPDITSYNDYFSL
: | : | | | | | : : : : | | | : : : : | | : : | | : : | | : | | : | |
VIFVQQALRKIPVQYAKRLVGRNPGVGQSTHLPLKVNAGVIPVIFALSLLIFPPTVAGLFGSDHPVAAWVIETFDY---
240 250 260 270 280 290 300

1173 1203 1233 1263 1293 1323 1353 1383
 SSIQGVVIYMILMLVLSVAFTFVNIDPTKISEAMRESGDFIPNYRPGKETQSYLSKICYLFGTFSGFFMAFLGGVPLLFA
 : : | : : | : : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
 THLIGMAVYALRIIGFTYFYAFIQVNPERMAENLKKQGGYIPGIRPGKATQTYITPILYRLTFVGSFLAVVAAILPVFF-
 320 330 340 350 360 370 380

1413 1440 1470 1500 1530 1560 1590 1620
LGNDLRTVSSMTGI-FMMITGMSFMILDEQVIRIRKQYTSVFENEEN*CFILFHLGIMKIVLGMIIITCGISSRLMSV
: || : | :: |:: :: : |:: |
IKFADLPQAIQIGGTGLLIVVGVALDTMKQIEAQLIKRSYKGFIK
 400 410 420 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1436-

Example 1299

A DNA sequence (GBSx1377) was identified in *S.agalactiae* <SEQ ID 3989> which encodes the amino acid sequence <SEQ ID 3990>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3002(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF61315 GB:U96166 unknown"[Streptococcus cristatus]

Identities = 30/78 (38%), Positives = 41/78 (52%)

Query: 276 ALTVTLTDDIWELEHLLQRCPNITDFHIAAPVYCSDRKQLVGYPNYLHEAITEEQFEVL 335

AL +T +D + ++E LL + PN FHI A S L L YPN L+ I + L

Sbjct: 289 ALILTNSDQLEQIEQLLTQLPNVHFHIGAITEMSGHLMGLNRYPNVSLYPNIRPAKVAEL 348

Query: 336 LLNSDIYLDINHGEVWN 353

D+YLDIN +E+ N

Sbjct: 349 FERCDLYLDINISDEILN 366

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1300

A DNA sequence (GBSx1378) was identified in *S.agalactiae* <SEQ ID 3991> which encodes the amino acid sequence <SEQ ID 3992>. This protein is predicted to be eps7. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC07458 GB:AX009404 product = eps7 [Streptococcus thermophilus]

Identities = 87/232 (37%), Positives = 133/232 (56%), Gaps = 22/232 (9%)

Query: 10 VSVIIPVYNAAPYLEGCVNTILGQTYQVFEILLIDGSTDTSASICDQLSLRDNRI RVFH 69

+S++IPVYN Y++ C+++IL QT+ EI+L+DDGSTD S ICD S D RI+V H

Sbjct: 3 ISIVIPVYNVQDYIKKCLDSILSQTFSDEILVDDGSTDLSGRICDYSENDKRIKVIH 62

Query: 70 IENGASKARNFGLARISPESQVTFVDSDDWVKENYLEVLLAQOEKYNADIVISNYIY 129

NGG S+ARN G+ + S+++TF+DSDD+V +Y+E L + +NADI I+++

Sbjct: 63 TANGGQSEARNVGIKNAT--SEWITFIDSDDYVSSDYIEYLYNLIQVHNADISIASF--- 117

Query: 130 RETEDIFGYIYTDKDFV-----IEEISAQTAIDRQVHWHLNSSVFIVIWGKLYRRELFD 183

YIT K + + + A+TAI R + LN + +WGK+YR E F+

Sbjct: 118 -----TYITPKKIIKHGNGEVALMDAKTAIRRL---LNEGFDMGVWGKMYRTEYFN 166

Query: 184 TITFPIDKVFEDLVSVLLFIKSKKTILVNGSYYGYRIRPNSIMTSAFSSKR 235

F K+FED L++ +F ++ + Y Y R NS + F+ K+

-1437-

Sbjct: 167 KYKFVSGKLFEDSLITYQIFSEASTIVFGAKDIYFVNRKNSTVNGTFNIIK 218

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1301

A DNA sequence (GBSx1379) was identified in *S.agalactiae* <SEQ ID 3993> which encodes the amino acid sequence <SEQ ID 3994>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1569(Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000(Not Clear) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 1302

A DNA sequence (GBSx1380) was identified in *S.agalactiae* <SEQ ID 3995> which encodes the amino acid sequence <SEQ ID 3996>. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.1662(Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000(Not Clear) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 1303

A DNA sequence (GBSx1381) was identified in *S.agalactiae* <SEQ ID 3997> which encodes the amino acid sequence <SEQ ID 3998>. This protein is predicted to be a glycosyl transferase (gspA). Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.2606(Affirmative) < succ>
bacterial membrane ---	Certainty=0.0000(Not Clear) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

-1438-

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
ducreyil]
Identities = 62/177 (35%), Positives = 105/177 (59%), Gaps = 8/177 (4%)

5 Query: 3 YARYYIPQLIDAEKVLVLDIDTLVVDNLDKLFIEELGDYPIAAILD--GDGIY-----FN 55
+ RY+I I+ +KV+YLD D +V +L +L++ ++ +Y +AA+ D + IY FN
Sbjct: 89 FFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISNYFLAAVKDIISEKIYVNNHIFN 148

10 Query: 56 SGVMLINSLYWMRYRVTEKLEITERELDNGIFGDQGVNLNLLFDNNWLKLEDKYNQVGN 115
+G++LIN+ W + +T+ L ++E+ +++ DQ +LNL+F + WLKL YN +G
Sbjct: 149 AGMLLINNKKWREHNITQFCLSLSEKYINSLPADQSIILNLFKDKWLKLNRGYNYLIGT 208

Query: 116 DLGAFYENWQGYFDRNFES-PTIIHYCTHDKPWNTFSSSRFRETWWQYEQLDWNEVF 171
D F Y + E+ P I IHY T KPW ++RFR +W Y +L+W +++
15 Sbjct: 209 DYLFCKYKTRYLEDLGETIPLIIHYNTEAKPWLNIENFRFRNIYWFYELNWQDIY 265

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 1304

A DNA sequence (GBSx1384) was identified in *S.galactiae* <SEQ ID 3999> which encodes the amino acid sequence <SEQ ID 4000>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

Possible site: 56
25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1157(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
ducreyil]
35 Identities = 103/259 (39%), Positives = 156/259 (59%), Gaps = 3/259 (1%)

Query: 7 IALAADFGYQEQVKTIKSIKSFHNFIDFYILNDDFPVVEWFQMMYHLSKMDCTISNTKI 66
I LAA+ Y E + T IKSI HN+ I FY+LN D+P EWF ++ L K++ I + K+
Sbjct: 10 IVLAANQSYSEYILTITIKSIYLNKHIRFYLLNRDYPTEWFDILNNKLRKLNSEIIDIKV 69

40 Query: 67 FNEEIKHFK-FQKMPYPTYFRYFIPEVIHEDKVLVLDLDCMIITSDLTSTFTDISKYGV 125
N+ IK+FK + T+FRYFI + I +DKV+YLD D+++ LT ++ DIS Y +
Sbjct: 70 TNDTIKNFKTYSHISSDTTFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISNYFL 129

45 Query: 126 AAVRDDLEEYDGKEDYFNSGLLLINNIFWREQGISQRLLDYTRENGALQYHDQDVLND 185
AAV+D + E+ FN+G+LLINN WRE I+Q L + + +L DQ +LN
Sbjct: 130 AAVKDIISEKIYVNNHIFNAGMLLINNKKWREHNITQFCLSLSEKYINSLPADQSIILNL 189

50 Query: 186 VLCDNWLELDETYNYHTGADMLYNLFQQSERQLNRRKDLPKVIHY-TATKPKWYLETSVR 244
+ D WL+L+ YNY G D L+ + ++ + + +P +IHY T KPW + + R
Sbjct: 190 IFKDKWLKLNRGYNYLIGTDYLFCKYKTRYLEDLGETIPLIIHYNTEAKPWLNI-FNTR 248

Query: 245 WRDIWWEYNRLEWRDIFTR 263
+R+I+W Y L W+DI+ +
55 Sbjct: 249 FRNIYWFYELNWQDIYAK 267

No corresponding DNA sequence was identified in *S.pyogenes*.

-1439-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1305

A DNA sequence (GBSx1385) was identified in *S.agalactiae* <SEQ ID 4001> which encodes the amino acid sequence <SEQ ID 4002>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2679(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus ducreyi]

Identities = 94/263 (35%), Positives = 158/263 (59%), Gaps = 4/263 (1%)

Query: 2 KKTIVLGADFQYRDQVMTTIKSIVSHNQHLTIYIINTDFPVEWFNINLNSLEQFDCRVKN 61

K IVL A+ Y + ++TTIKSI HN+H+ Y++N D+P EWF+ILN+ L + + + +

Sbjct: 7 KMNIVLAANQSYSEYILTITIKSIYLNKHIRFYLLNRDYPTEWFDILNNKLRLNSEIID 66

Query: 62 IPISSDVFEIGIPTLSHISV-AGFFRWFPIHLEEEIVLYLSDVIVRGSGLDPLFDINLEE 120

I +++D + T SHIS FFR+FI +E++ V+YLD+D++V GSI L+ ++

Sbjct: 67 IKVTNDTIKNFKTYSHISSDTTFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISN 126

Query: 121 NLLGAVADHFSTLYYGDTPVVSFNSGVMLINNSLWKKEIYNLSLMRIADKG-SAVGVGDQ 179

L AV D S Y + FN+G++LINN W++ I + +++K +++ DQ

Sbjct: 127 YFLAAVKDIISEKIYVNNH--IFNAGMLLINNKKWREHNITQFCLSLSEKYINSLPDADQ 184

Query: 180 EYLNILITQNRWIDIGKQYNVQIGQDVNINAYGRPDLYHFYDDCEPVIVHYNSQDKPWPNKY 239

LN++ +++W+ + + YN IG D YG+ + P+I+HYN++ KPW

Sbjct: 185 SILNLIIFKDKWLKLNRGYNYLIGTDYLFYKYGKTRYLEDLGETIPLIIHYNTEAKPWLNI 244

Query: 240 SQSRYRSEWYFGLWESVIYAQ 262

+R+R+ +W+Y+ L W IYA+

Sbjct: 245 FNTFRFNIIYWFYELNWQDIYAK 267

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1306

A DNA sequence (GBSx1386) was identified in *S.agalactiae* <SEQ ID 4003> which encodes the amino acid sequence <SEQ ID 4004>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2996(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1440-

A related GBS nucleic acid sequence <SEQ ID 10003> which encodes amino acid sequence <SEQ ID 10004> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAC75095 GB:AE000294 putative Galf transferase [Escherichia coli K12]
    Identities = 68/286 (23%), Positives = 122/286 (41%), Gaps = 18/286 (6%)

Query: 77 STRMDGIIAGLGRGDIVVFQVPTWNSTEFDELFLDKLQAYGARIITTFVHDI VPLMFESNF 136
      S ++ + GL D+++F P F +L + RI+ +HDI L
Sbjct: 50 SVKLSTFLCGLNKDVLIFNFPMAKPFWHILSFFHRLKLF--RIVPLIHDIDELRGGGGS 107

10 Query: 137 YLLDRVIDMYNRSDDVILPTKAMHDYLIIEKGMTTSKVLVQEVWDHPVNIDLP RPEC---Q 193
      D V D+VI M YL K M+ K+ +++D+ V+ D+ + Q
Sbjct: 108 ---DSV--RLATCDMVISHNPQMTKYL-SKYMSQDKIKDIKIFDYLVSSEVHRDVTDKQ 161

15 Query: 194 KVL SFAGDIQRFPFVNDWKENIPLIYYGDSRLNSEANVHAQGWKDDVELMLSLSKRG-G 252
      + + +AG++ R + E +G ++ N G D + ++ G
Sbjct: 162 RGVIIYAGNLSRHKCSFTYEGCDFTLFG--VNYENKDNPKYLG-SFDAQSPEKINLPGMQ 218

20 Query: 253 FGLCWESEDREELVERR---YSRMNASYKLSTFLAAGLPIIANHDISSRDFIKQHGLGFTV 309
      FGL W D E Y + N +K S +L+ LP+ + DFI + +G+ V
Sbjct: 219 FGLIWDGDSVETCSGAFGDYLFKNPNPKTSLYLSMELPVFIWDKAALADFI VDNRI GYAV 278

Query: 310 ETLEEAVEKINMEKETYSYVENVEKIATLLRNGYITKLLIDAV 355
      +++E E +++M ETY EN + I+ +R G + +L + +
25 Sbjct: 279 GSIKEMQEIVDSMTIETKYQISENTKIISQKIRTGSYFRDVLEEVI 324

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1307

A DNA sequence (GBSx1387) was identified in *S.agalactiae* <SEQ ID 4005> which encodes the amino acid sequence <SEQ ID 4006>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3098(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA73093 GB:M76233 [Rabbit smooth muscle myosin light chain
    kinase mRNA, complete CDS.], gene product [Oryctolagus
    cuniculus]
45 Identities = 23/63 (36%), Positives = 36/63 (56%)

Query: 5 QPAPALQVRVRCQPAPVLQPVPRCQPALALQVRVRCQPAQVLQQVPRCQPAQVLQQVPRC 64
      +PA L+ V +PA L+PV +PA L+ V +PA+ L+ V +PA+ L+ V
Sbjct: 225 KPAETLKPVGNAKPAETLKPVGNAKPAETLKPVGNAKPAETLKPVGNAKPAETLKAVANA 284

50 Query: 65 QPA 67
      +PA
Sbjct: 285 KPA 287

```

55 No corresponding DNA sequence was identified in *S.pyogenes*.

-1441-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1308

A DNA sequence (GBSx1388) was identified in *S.agalactiae* <SEQ ID 4007> which encodes the amino acid sequence <SEQ ID 4008>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.24	Transmembrane	189 - 205 (173 - 245)
INTEGRAL	Likelihood = -9.24	Transmembrane	213 - 229 (206 - 245)
INTEGRAL	Likelihood = -7.96	Transmembrane	95 - 111 (83 - 185)
INTEGRAL	Likelihood = -7.96	Transmembrane	115 - 131 (112 - 185)
INTEGRAL	Likelihood = -7.96	Transmembrane	135 - 151 (132 - 185)
INTEGRAL	Likelihood = -7.96	Transmembrane	155 - 171 (152 - 185)
INTEGRAL	Likelihood = -6.85	Transmembrane	15 - 31 (8 - 45)
INTEGRAL	Likelihood = -4.09	Transmembrane	39 - 55 (35 - 57)
INTEGRAL	Likelihood = -4.09	Transmembrane	63 - 79 (59 - 81)
INTEGRAL	Likelihood = -2.71	Transmembrane	235 - 251 (235 - 251)
INTEGRAL	Likelihood = -0.11	Transmembrane	253 - 269 (253 - 269)

----- Final Results -----

bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC16164 GB:AF010496 ice nucleation protein [Rhodobacter capsulatus]
Identities = 85/286 (29%), Positives = 119/286 (40%), Gaps = 17/286 (5%)

Query: 3 ALVLADVDALVETLVLADVVALLIEALVLADIEALV----EALVLADIEALVEALVLADID 58
AL A AL T + A ++ L AD+ L +AL A I AL + + A
Sbjct: 523 ALSDAQAGALTSTQIGLLSTA AVKGLSTADMACLT TAEQAALTSQAIAALSSSQIRAMTT 582

Query: 59 ALVEALVLADIEALVEALVL----ADIDALVEALVLADVEALIEALVLALVEALVLADVE 114
A + AL A I+ L + +L ADI AL A + I AL +LV A+ AD+
Sbjct: 583 AQIAALGTAQIKGLTASNILGLETADIVALT TQAPALSSSQIAALSTSLVAAMETADLA 642

Query: 115 ALIEALVLAL----VEALVLADVEAL----IEALVLALVEALVLADVEALIEALVLALVE 166
L A + AL A A+ I + A ++ L AD+ AL A + +
Sbjct: 643 KLSAATFKGFSSTQITALT TTAQAGAIGTDQIAQITTA AIKGLSADIAALANATLAKMTT 702

Query: 167 ALVLADVEALIEALVLADVD-ALVLALVEALVLALVEALILAEVEALVLALVEALVLALV 225
A V A + L ++ L A V+AL A + L ++ AL AL V
Sbjct: 703 AQVAVLGSQTLGTLTTTQINTVLTTAQVKALGAAALAGLGTDDIVALT TQAAAALSSSTQV 762

Query: 226 EALILALVEALVLADVDALMEALVLADVEALMEALVLADVDALVEA 271
AL A + AL AD AL A + + AL +DAL A
Sbjct: 763 AALSTAQISALQTADFAALSTA AIKGLSSTQITALTSTGQIDALTTA 808

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1309

A DNA sequence (GBSx1389) was identified in *S.agalactiae* <SEQ ID 4009> which encodes the amino acid sequence <SEQ ID 4010>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

-1442-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2297(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1310

A DNA sequence (GBSx1390) was identified in *S.galactiae* <SEQ ID 4011> which encodes the amino acid sequence <SEQ ID 4012>. This protein is predicted to be fimbriae-associated protein Fap1. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA97453 GB:AB029393 streptococcal hemagglutinin [Streptococcus gordonii]
 Identities = 388/968 (40%), Positives = 518/968 (53%), Gaps = 68/968 (7%)

Query: 13 VDTKSRVIMHKSEKNWVRTVMSHFNLFKAIKGRATVEADVCIQDVEKEDRLSSGNLTYLK 72
 V+ +R K+ KS K+W+R S F L + +KG +V V +E + G L YLK
 Sbjct: 13 VERVTRFKLIKSGKHWLRAATSQFGLRLMKGADISSVEV---KVAEEQSVKGGGLNYLK 69

Query: 73 GILAAGALVGGASLTSR-VYADETPVQEQSSSVPTLAEQTEVTV--KTTTVQNHQDGTV 129
 GI+A GA++GGA +TS VYA+E +++ + LA + E + + T + +
 Sbjct: 70 GIIATGAVLGGAVVTSSSVYAEQEALEKVIDTRDVLATRGEAVLSEEAATLSSEGANP 129

Query: 130 SKNIIDSNSVSMSESASTSTSESVMMSGSTLTSVSESVSTSALTSESISTSESSES 189
 +++ D+ S S S SA+ S S S+S+S S S S S S S+S S+SES S S S SV
 Sbjct: 130 VESLSDTLSASESASAN-SVSTSISSISFSVSASASLSSSSLSQSSSESASASELSV 188

Query: 190 SKSTSISEVSNILETQASLTDKGRESFANQIVTESSLVTDAGKNASVSSLIEITKPKSE 249
 S STS S S TQ+S + S S+N + T S V+ +NA V + + +E
 Sbjct: 189 SASTSQSFSSTTSSTQSSNNESLISSDSSNSLNTNQS-VSARNQARVTRRAVAANDTE 247

Query: 250 LQTSKMSNESLITPEKSQVMIASDKTGNESLTPTRLKSVIQPRSMNLMTLSSEMDLIPL 309
 K + + E + ++ T N + ++ N+ ++ L P
 Sbjct: 248 APQVKSGDYVVYRGESFEYY--AEITDMSGQVNRVIR-----NVEGGANSTYLSNP 297

Query: 310 EEVSDTEMLGKDVSSSELQKVNIALKDNTLSEPGTVKLDSSENLVNFAFSIASVNEGDVF 369
 TE LG+ ++ +Q L+ E ++ + ++ + +A G+
 Sbjct: 298 WVKYSTENLGRPGNATVQN---PLRTRIPGEVPLNEIVNEKSYTRYI--VAWDPSGN-- 350

Query: 370 TVKLSDNLDTOGIGTILKVQDIMDETQQLLATGSYSPLTHNITY-----TWTRYAST 421
 ++ DN + G+ + +E Y P ++TY T R A
 Sbjct: 351 ATQMVNDNANRNGLERFVLTVKSQNE-----KYDPAESSVTYVNNLSNLSSTSEREAVA 402

Query: 422 LNNIKARVNMPVWPDQRI-----ISKTTSDKQCFTATLNNQVASIE---ERVQYNSPS 471
 A N+P P +I ++ T DK T N V ++ S S
 Sbjct: 403 AAVRAANPNIP--PTAKITVSQNGTVTITTPDKSTDTIPANRVVKDLQISKSNSASQSSS 460

Query: 472 VTEHTNVKTNVRSRIMKLDDERQTETYITQINPEGKEMYFASGLGNLYTIIGSDGTSQSP 531

-1443-

V+ + T+V + I ++ + + ++ S+ S S
 Sbjct: 461 VSASQASASTSVSASI---SASMSASVSVSTASASTSASVSASESASTSASVSASESASTS- 516
 Query: 532 VNILNAEVKILKTNSKNLTDSDQNYDSPEFEDVTSQYSYTNDSKITTIDWKTNISISSTT 591
 A V K++S + + S ++ + + S + S + S+S++T
 Sbjct: 517 -----ASVSASKSSSTASVSASESASTSASVSASESASTSASVSASESASTSASVSAST 571
 Query: 592 SYVVLVKIPKQSGVLYSTVSDINQTYGSKYSYGHNTNISGDSANAEIKL-LSESASTSAS 650
 S + ST + ++ + + S ++S A+ + SESASTSAS
 Sbjct: 572 SASTSASVSASESA--STSASVSASESASTS--ASVSASESASTSASVSASESASTSAS 626
 Query: 651 TSASTSASMSASTSASTSASMSASTSASTSASTSASMSASTSASTSASTSASTSASTSAS 710
 SAS S+S SAS SAS SAS SAS SAS SASTSAS+SASTSASTSAS SASTSASTSAS
 Sbjct: 627 VSASESSSTASVSASESASTSASVSASESASTSASVSASTSASTSASTSASTSASTSAS 686
 Query: 711 MSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 770
 +SASTSASTSAS SAS SASTSAS SAS SASTSAS SASTSASTSAS+SASTSASTSAS
 Sbjct: 687 VSASTSASTSASVSASESASTSASVSASESASTSASVSASTSASTSASTSASTSASTSAS 746
 Query: 771 TSASTSASMSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 830
 SAS SAS SAS SASTSASTSAS SAS SASTSAS SAST ASTSAS+SAS SASTSAS
 Sbjct: 747 VSASESASTSASVSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 806
 Query: 831 TSASMSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 890
 SAS SASTSAS SAS SASTSAS SAS SASTSAS SAS SASTSAS SAS SASTSAS
 Sbjct: 807 VSASESASTSASVSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 866
 Query: 891 MSATTSTASTSVSTASTSASTSASTSASTSSSSSVTSNSSKEKVYALPSTGDDQDYSVTATLALG 950
 +SA+TSASTS S SAS SASTSAS S+S S ++++S SA S +T+
 Sbjct: 867 VSASTSASTSASVSASESASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 926
 Query: 951 LGLMTGAT 958
 + T A+
 Sbjct: 927 VSASTSAS 934

There is also homology to SEQ ID 760.

SEQ ID 4012 (GBS68) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 4; MW 131.2kDa).

GBS68d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 153 (lane 14; MW 103kDa) and in Figure 239 (lane 13; MW 103kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 17; MW 78kDa), in Figure 153 (lane 17; MW >78kDa) and in Figure 184 (lane 10; MW 78kDa). Purified GBS68d-GST is shown in Figure 246, lane 5.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1311

A DNA sequence (GBSx1391) was identified in *S.agalactiae* <SEQ ID 4013> which encodes the amino acid sequence <SEQ ID 4014>. This protein is predicted to be RofA. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1444-

A related GBS nucleic acid sequence <SEQ ID 10005> which encodes amino acid sequence <SEQ ID 10006> was also identified.

There is also homology to SEQ ID 3750.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1312

A DNA sequence (GBSx1392) was identified in *S.agalactiae* <SEQ ID 4015> which encodes the amino acid sequence <SEQ ID 4016>. This protein is predicted to be Nra. Analysis of this protein sequence
10 reveals the following:

Possible site: 16
>>> Seems to have a cleavable N-term signal seq.

15 ----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

There is also homology to SEQ ID 3750.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1313

A DNA sequence (GBSx1393) was identified in *S.agalactiae* <SEQ ID 4017> which encodes the amino acid sequence <SEQ ID 4018>. Analysis of this protein sequence reveals the following:

25 Possible site: 19
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3674 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAA27020 GB:M80215 uvs402 protein [Streptococcus pneumoniae]
Identities = 577/663 (87%), Positives = 633/663 (95%), Gaps = 1/663 (0%)

Query: 1 MIDRKDTNRFLVSKYSPSGDQPQAIETLVDNIEGGEKAQILKGATGTGKTYTMSQVIAQ 60
MI+ N+FKLVSKY PSGDQPQAIETLVDNIEGGEKAQIL GATGTGKTYTMSQVI++
Sbjct: 7 MINHITDNQFKLVSKYQPSGDQPQAIETLVDNIEGGEKAQILMGATGTGKTYTMSQVISK 66

40 Query: 61 VNKP TLVIAHNKTLAQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV 120
VNKP TLVIAHNKTLAQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV
Sbjct: 67 VNKP TLVIAHNKTLAQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV 126

45 Query: 121 NDEIDKL R HSATSSLLERN DVIVVASVSCIYGLGSPKEYADSVVSLRPGQEISRDL LN 180
NDEIDKL R HSATSSLLERN DVIVVASVSCIYGLGSPKEYADSVVSLRPGQEISRDL LN
Sbjct: 127 NDEIDKL R HSATSSLLERN DVIVVASVSCIYGLGSPKEYADSVVSLRPGQEISRDL LN 186

50 Query: 181 LVDIQFERNDIDFQRGKFRVRGDDVVEFPASRDEHAFRIEFFGDEIDRIREIESLTGRVL 240
LVDIQFERNDIDFQRGKFRVRGDDVVEFPASRDEHAFRIEFFGDEIDRIREIESLTGRVL
Sbjct: 187 LVDIQFERNDIDFQRGKFRVRGDDVVEFPASRDEHAFRIEFFGDEIDRIREIESLTGRVL 246

-1445-

Query: 241 GEVEHLAIFPATHFMTNDEHMEEAISKIQAEMENQVELFEKEGKLIEAQRIRQRTEYDIE 300
 GEV+HLAIFPATHF+TND+HME AI+KIQAE+E Q+ +FEKEGKL+EAQR++QRTEYDIE
 Sbjct: 247 GEVDHLAIFPATHFVTNDDHMEVAIAKIQAELEEQLAVFEKEGKLLEAQRRLQRTEYDIE 306

5 Query: 301 MLREMGYTNGVENYSRHMDGRSEGEPPFTLLDFFPEDFLIMIDESHMTMGQIKGMYNGDR 360
 MLREMGYTNGVENYSRHMDGRSEGEPP+TLLDFFP+DFLIMIDESHMTMGQIKGMYNGDR
 Sbjct: 307 MLREMGYTNGVENYSRHMDGRSEGEPPYTLDDFFPDDFLIMIDESHMTMGQIKGMYNGDR 366

10 Query: 361 SRKEMLVNYGFRLPALSALDNRPLRREEFESHVHQIVYVSATPGDYEMEQTDFTVVEQIIRPT 420
 SRK+MLVNYGFRLPALSALDNRPLRREEFESHVHQIVYVSATPGDYE EQT+TV+EQIIRPT
 Sbjct: 367 SRKKMLVNYGFRLPALSALDNRPLRREEFESHVHQIVYVSATPGDYENEQTETVIEQIIRPT 426

15 Query: 421 GLLDPEVEVRPSMGQMDLLGEINLRTEKGERTFITTLTKRMAEDLTDYKEMGVKVKYM 480
 GLLDPEVEVRP+MGQ+DDLGEIN R EK ERTFITTLTK+MAEDLTDY KEMG+KVKYM
 Sbjct: 427 GLLDPEVEVRPTMGQIDDLGEINARVEKNERTFITTLTKKMAEDLTDYFKEMGIKVKYM 486

20 Query: 481 HSDIKTLERTEIIRDLRLGVFDVLIGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI 540
 HSDIKTLERTEIIRDLRLGVFDVL+GINLLREGIDVPEVSLVAILDADKEGFLRNERGLI
 Sbjct: 487 HSDIKTLERTEIIRDLRLGVFDVLVGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI 546

25 Query: 541 QTIGRAARNSNGHVIMYADKITDSMQRAMDETARRRRQMODYNEKHGIVPQTIKKEIRDL 600
 QTIGRAARNS GHVIMYAD +T SMQRA+DETARRR++QM YNE+HGIVPQTIKKEIRDL
 Sbjct: 547 QTIGRAARNSEGHVIMYADTVTQSMQRAIDETARRRRQIMAYNEEHGIVPQTIKKEIRDL 606

30 Query: 601 IAITKSNDSKPEKVVDYSSLSKKERQAEIKALQQQMGEAEELDFELAAQIRDVILELK 660
 IA+TK+ ++ +K VD +SL+K+ER+ +K L++QMGEA E+LDFELAAQIRD++LE+K
 Sbjct: 607 IAVTKAVAKEE-DKEVDINSLNQERKELVKKLEKQMGEAVEVLDLDFELAAQIRDMMLEVK 665

Query: 661 AID 663
 A+D
 Sbjct: 666 ALD 668

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4019> which encodes the amino acid sequence <SEQ ID 4020>. Analysis of this protein sequence reveals the following:

35 Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.4386(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 570/663 (85%), Positives = 625/663 (93%)

45 Query: 1 MIDRKDTNRFKLVSKYSPSGDQPQAIETLVDNIEGGEKAQILKGATGTGKTYTMSQVIAQ 60
 MID++D FKL SKY PSGDQPQAIETLVDNIEGGEKAQIL GATGTGKTYTMSQVI++
 Sbjct: 1 MIDKRDDKPFKLKSKYKPSGDQPQAIETLVDNIEGGEKAQILGATGTGKTYTMSQVISK 60

50 Query: 61 VNKPTLVIAHNKTLAGQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV 120
 VNKPTLVIAHNKTLAGQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV
 Sbjct: 61 VNKPTLVIAHNKTLAGQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV 120

55 Query: 121 NDEIDKLRSATSSLLERNDVIVVASVSCIYGLGSPKEYADSVSLRPGQEISRDQLNN 180
 NDEIDKLRSATSSLLERNDVIVVASVSCIYGLGSPKEYADS VSLRPGQEISRD LLN
 Sbjct: 121 NDEIDKLRSATSSLLERNDVIVVASVSCIYGLGSPKEYADSAVSLRPGQEISRDTLNNQ 180

60 Query: 181 LVDIQFERNDIDFQRGKFRVRGVDVVEVFPASRDEHAFFRIFEGDEIDRIRIESLTGRVL 240
 LVDIQFERNDIDFQRG FRVRGVDVVEVFPASRDEHAFFR+EFFGDEIDRI EIESLTG+ +
 Sbjct: 181 LVDIQFERNDIDFQRCFRVRGVDVVEVFPASRDEHAFFRVEFFGDEIDRICEIESLTGKTI 240

65 Query: 241 GEVEHLAIFPATHFMTNDEHMEEAISKIQAEMENQVELFEKEGKLIEAQRIRQRTEYDIE 300
 GEV+HL +FPATHF+TNDHME++I+KIQAE+ Q++LFE EGKL+EAQR+RQRTEYDIE
 Sbjct: 241 GEVDHLVLFPAHFVTNDEHMEQSIKIQAEAEQLQLFESEGLLEAQRRLQRTEYDIE 300

-1446-

Query: 301 MLREMGYTNGVENYSRHM DGRSEGEPPFTLLDFFPEDFLIMIDESHMTMGQIKGMYNGDR 360
MLREMGYT+GVENYSRHM DGRS GEPP+TLLDFFPEDFLIMIDESHMTMGQIKGMYNGD+
Sbjct: 301 MLREMGYTSGVENYSRHM DGRSPGEPPYTL DFFPEDFLIMIDESHMTMGQIKGMYNGDQ 360

5 Query: 361 SRKEMLVNYGFR LPSALDNRPLRREEFESHVHQIVVVSATPGDYEMEQTDTVVEQIIRPT 420
+RK+MLV+YGFR LPSALDNRPLRREEFESHVHQIVVVSATPG+YEM QT+T++EQIIRPT
Sbjct: 361 ARKQMLVDYGFRLPSALDNRPLRREEFESHVHQIVVVSATPGYEMSQTNTTIEQIIRPT 420

10 Query: 421 GLLDPEVEVRPSMGQMDLLGEINLRTEKGERTFITTLTKRMAEDLTDY LKEMGVKVKYM 480
GLLDPE++VR SMGQMDLLGEIN R + ERTFITTLTK+MAEDLTDY LKEMGVKVKYM
Sbjct: 421 GLLDPEIDVRSSMGQMDLLGEINQRVARDERTFITTLTKRMAEDLTDY LKEMGVKVKYM 480

15 Query: 481 HSDIKTLERTEIIRDLRLGVFDVLIGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI 540
HSDIKTLERTEIIRDLRLGVFDVLIGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI
Sbjct: 481 HSDIKTLERTEIIRDLRLGVFDVLIGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI 540

20 Query: 541 QTIGRAARNSNGHVIMYADKI TDSMQRAMDETARRRLQMDYNEKHGIVPQTIKKEIRDL 600
QTIGRAARN +GHVIMYADK+TDSMQRA+DETARRR +Q+ YN+ HGIVPQTIKK+IR L
Sbjct: 541 QTIGRAARNVDGHVIMYADKMTDSMQRAIDETARRREIQIAYNKAHGIVPQTIKKDIRGL 600

25 Query: 601 IAITKSNDSDKPEKVVDYSSLSKKERQAEIKALQQMQEAAELLD FELAAQIRDVILELK 660
I+I+K++ +D ++ +DY S+S+ ER+ I ALQ+QM QEAAELLD FELAAQ+RD+ILELK
Sbjct: 601 ISISKTSHNDISKEEMDYESMSRGERKEAINALQKMQEAAELLD FELAAQMRDLILELK 660

Query: 661 AID 663
+D
Sbjct: 661 LMD 663

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

Example 1314

A DNA sequence (GBSx1394) was identified in *S. agalactiae* <SEQ ID 4021> which encodes the amino acid sequence <SEQ ID 4022>. Analysis of this protein sequence reveals the following:

Possible site: 31

35 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.78	Transmembrane	284 - 300 (274 - 303)
INTEGRAL	Likelihood = -10.08	Transmembrane	20 - 36 (16 - 53)
INTEGRAL	Likelihood = -5.52	Transmembrane	117 - 133 (114 - 137)
INTEGRAL	Likelihood = -5.15	Transmembrane	203 - 219 (201 - 225)
40 INTEGRAL	Likelihood = -3.29	Transmembrane	183 - 199 (182 - 200)
INTEGRAL	Likelihood = -1.54	Transmembrane	74 - 90 (73 - 90)
INTEGRAL	Likelihood = -0.48	Transmembrane	37 - 53 (37 - 53)

----- Final Results -----

45 bacterial membrane --- Certainty=0.5713(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:CAA22372 GB:AL034446 putative transmembrane protein
[Streptomyces coelicolor A3(2)]
Identities = 58/190 (30%), Positives = 96/190 (50%), Gaps = 11/190 (5%)

Query: 114 GWS--IGFILFSISVITAYILGG LDFHSYDVS K-ATIFYVV TLLPFWLIQSGTEELLTRG 170
GW IGF LF +VIT G Y+V ++ + L+ F + TEE++ RG
55 Sbjct: 98 GWGTLIGFGLFG-AVITNL FASGY----YEVDGLGSVQGAIGLVGFMAAAAATEEVVFRG 152

Query: 171 WLLPLINHRFHLAVAIGVSSTLFGILHLVNAHVTFLSIVSI-ICSGVLSLYMIKSGNIW 229
L +I +A+G++ +FG++HL+N T ++I I +G +++ + N+W
60 Sbjct: 153 VLFRIIEEHIGTYLALGLTGLVFGMLHLNEDATLWGALAIAT EAGFMLAAAYAATRN LW 212

Query: 230 SVAALHGAWNFSQGNLYGIAVSGQKAGASLLHFTV KENAPDWISGGAFGIEGSLISIFVL 289

-1447-

+H WNF+ G ++ VSG LL T+ + P ++GG FG EGS+ S+
 Sbjct: 213 LTIGVHFGWNFAAGGVFSTVVSNGDSEGLLDATM--SGPKLLTGDFGPEGSVYSVGFG 270
 Query: 290 LAATYYLLWL 299
 + + LWL
 Sbjct: 271 VLLTLVFLWL 280

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1315

A DNA sequence (GBSx1395) was identified in *S.galactiae* <SEQ ID 4023> which encodes the amino acid sequence <SEQ ID 4024>. This protein is predicted to be glutamine-binding periplasmic protein/glutamine transport system perme. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -8.97 Transmembrane 532 - 548 (523 - 553)
 INTEGRAL Likelihood = -7.38 Transmembrane 700 - 716 (696 - 720)
 INTEGRAL Likelihood = -4.57 Transmembrane 562 - 578 (558 - 588)
 INTEGRAL Likelihood = -0.32 Transmembrane 665 - 681 (665 - 681)

----- Final Results -----

bacterial membrane --- Certainty=0.4588(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF16724 GB:AF141644 putative integral membrane protein
 [Lactococcus lactis]

Identities = 109/195 (55%), Positives = 156/195 (79%), Gaps = 4/195 (2%)

Query: 466 KMFNNGLASLKKSGEYDKLVKKYLSTASTSSNDKAAKPVDESTILGLISNNYKQLLSGIG 525
 +MFNNGLA+L+ +GEYDK++ KYL++ T + +AK E+T G++ NN++Q+ G+
 Sbjct: 1 EMFNNGLANLRANGEYDKIIDKYLAS-DTKTIQSSAK---ENTFFGILQNNWEQIGRGLL 56

Query: 526 TTLSLTLSFAIAMVIGIIFGMMSVSPSNTLRTISMIFVDIVRGIPLMIVAFAIFWGIPN 585
 TL L ++SF +AM++GIIFG+ SV+PS LRTI+ I+VD+ R IPL+++ FIF+GIPN
 Sbjct: 57 VTLELAVLSFILAMIVGIIFGLFSVAPSKILRTIARIYVDLNRSIPLLVLTIFIFYGIPN 116

Query: 586 LIESITGHQSPINDFVAATIALSLNGGAYIAEIVRGGIEAVPSGQMEASRSLGISYGKTM 645
 L++ ITGHQSP+N+F A IAL+LN AYIAEIVR G++AVPSGQMEASRSLG++Y +M
 Sbjct: 117 LLQIITGHQSPLNFTAGVIALTLNSSAYIAEIVRSGVQAVPSGQMEASRSLGVTYLTSM 176

Query: 646 QKVILPQAVRLMLPN 660

+KVILPQA+++ +P+

Sbjct: 177 RKVILPQAIKITIPS 191

There is also homology to SEQ ID 1198.

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9071> which encodes amino acid sequence <SEQ ID 9072>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1448-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 80.8 bits (196), Expect = 2e-17
 Identities = 64/233 (27%), Positives = 113/233 (48%), Gaps = 13/233 (5%)

Query: 34 IKKTRKLVVAVSPDYAPFEFKALVNGKDTIVGADVQLAQAIADDELVDLELSPMSFDNVL 93
 +K + K+V S +APFE++ NGK G D++L + IA + L++S FD L
 Sbjct: 268 VKPSYKIVSDSS--FAPFEYQ---NGKGKYTGFDMELIKIAKQQGFKLDISNPGFDAAL 322

Query: 94 SSLQTKGADLAISGISHTKERAKVYDFSIPYYQAENAIVMRASDAKVTKNISDLNGKKVA 153
 +++Q+G+AD I+G + T+ R K++DFS PYY +++++ K+ DL GK V
 Sbjct: 323 NAVQSGQADGVIAGATITEARQKIFDFSDPYY--TSSVILAVKKGSNVKSQDLKGKTVG 380

Query: 154 AQKGSIEEGLVKIQLPKANLISLTAMGEA---INELKAGQVYAVTLEAPVAAGFLAQHKD 210
 A+ G+ + K N + A EA + + +G + A+ + V A + Q +
 Sbjct: 381 AKNGTASYTWLSDHADKYN-YHVKADEASTMYDSMNSGSIDALMDDEAVLAYAINQGRK 439

Query: 211 LALAPFSLKTS DGDAKVALPKNSGDLTKAVNKVIAKLDEQERYKSFI AETIA 263
 P + S GD + +L K N +A L + Y + + ++
 Sbjct: 440 FE-TPIKGEKS-GDIGFAVKKGANPELIKMFNGLASLKKSGEYDKLVKKYLS 490
 Score = 74.5 bits (180), Expect = 1e-15
 Identities = 59/215 (27%), Positives = 102/215 (47%), Gaps = 12/215 (5%)

Query: 48 YAPFEFKALVNGKDTIVGADVQLAQAIADDELVDLELSPMSFDNVLSLQTKGADLAISG 107
 YAPFEFK + T G DV + +A ++ ++ FD ++++Q+G+AD ++G
 Sbjct: 36 YAPFEFK---DSDQTYKGIDVDIVNEVAKRAGWNVNMTYPGFDAAVNAVQSGQADALMAG 92

Query: 108 ISHTKERAKVYDFSIPYYQAENAIVMRASDAKVTKNISDLNGKKVAAQKGSIEEGLVKIQ 167
 + T+ R KV++PS YY + I+ ++ KVT N L GK V + G+ + ++
 Sbjct: 93 TTVTEARKKVFNFSDTYYDT-SVILYTKNNNKVT-NYKQLKGKVVGVKNGTAAQS FLEEN 150

Query: 168 LPKANLISLTAMGEAI--NELKAGQVYAVTLEAPVAAGFLAQHKDLALAPFSLKTS DGDA 225
 K T + N L +G +YA + PV + Q K A+ +++ +
 Sbjct: 151 KSKYGYKVKTFDTSDLMNNSLDSGSIYAAMDDQPVVQFAINQGKAYAI---NMEGEAVGS 207

Query: 226 KVALPKNSG--DLTKAVNKVIAKLDEQERYKSFI 258
 A A+ K SG +L K N A++ Y +
 Sbjct: 208 FAFVKKKSGHDNLIKEFNTAFQMKSDGTYNDIM 242

SEQ ID 4024 (GBS154) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 199, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1316

A DNA sequence (GBSx1396) was identified in *S.agalactiae* <SEQ ID 4025> which encodes the amino acid sequence <SEQ ID 4026>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein (glnQ). Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4183(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB90561 GB:AE001058 glutamine ABC transporter, ATP-binding

-1449-

protein (glnQ) [Archaeoglobus fulgidus]
Identities = 147/240 (61%), Positives = 192/240 (79%)

5 Query: 5 KIDVQDLHKSQYQNEVLKGIDAKFYEGDVVCIIGPSGSGKSTFLRTLNLLESITSGKVVV 64
++++ DLHK +G+ EVLKG+ K +G+VV IIGPSGSGKST LR +N LE TSGK+++
Sbjct: 3 QLEIIDLHKRFGELEVLKGVMTKVEKGEVVVCIIGPSGSGKSTLLRCINRLEPTSGKILL 62

10 Query: 65 DGFELSNPKTDIDKARENIGMVFOHFNLFPHMSVLENIITFAPIELGKESKEAAEKHGMEL 124
DG +++N K DI+K R+ IG+VFQ FNLFPH++ L+N+T API++ K SK AE+ GM L
Sbjct: 63 DGVDITNSKIDINKVRQRIGIVFQQFNLFPHLTALQNVTLAPIKIKKMSKREAEELGMRL 122

15 Query: 125 LEKVGLADKANAKPDSLGGQKQRVAIARSLAMNPDILLFDEPTSAIDPEMVGDLNVLMK 184
LEKVGL DKA+ P LSGGQ+QRVAIAR+LAMNP+++LFDE TSALDPE+V +VL+VMK
Sbjct: 123 LEKVGLEDKADYYPAQLSGGQQQRVAIARALAMNPEVMLFDEVTSALDPELVKEVLDVMK 182

Query: 185 DLAEQGMTMLIVTHEMGFARQVANRVIFTDGGRFLEDGTPEQIFDTPQHPRLDQFLNKVL 244
LA GMTM++VTHEMGFAR+V +RVIF DGG +E+G PEQIF P+H R + FL+ +L
Sbjct: 183 QLARDGMTMVVVTHEMGFAREVGDRVIFMDGGVIVEEGKPEQIFSNPKHERTRKFLSMIL 242

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4027> which encodes the amino acid sequence <SEQ ID 4028>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4149(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the databases:

>GP:BAB05180 GB:AP001512 ABC transporter (substrate-binding protein)
[Bacillus halodurans]
Identities = 79/227 (34%), Positives = 126/227 (54%), Gaps = 10/227 (4%)

35 Query: 35 KKTRKLVAVSPDYAPFEFKALVNGKDTIVGADVQLAQAIADLVDLELSPMSFDNVLS 94
+K LV+ S DY P+E + G+ IVG DV +A+ I EL +L++ M F+ ++
Sbjct: 48 EKKSVLVMGTSADYPPYESVDVTTGE--IVGFDVDIAEYITSELGYELKIQDMDFNIGIIP 105

40 Query: 95 SLQTGKADLAISGISHTKERAKVYDFSIPYYQAENAIVMRASDAKVTKNISDLNGKKVAA 154
+LQ G+ D A+SG++ T+ER K DFS YY A+N +V + D ++ DL GK V
Sbjct: 106 ALQAGRVDFAISGMTPTTEERKKSVDVDFVYDAQNLVVFKEEDG--LSSVEDLAGKTVGV 163

45 Query: 155 QKGSII-EEGLVKIQ--LPKANLISLTAMGEAINELKAGQVYAVTLEAPVAAGFLAQHKDL 211
Q SI EE V++Q L + + + E + EL AG+V A+ +E VAAG L +
Sbjct: 164 QLASIQEEAAVELQEELDGLTIETRNRPVPELVQELLAGRVDALIIEDTVAAGHLEANP-- 221

Query: 212 ALAPFSLKTSDDAKAVALPKNSGDLTKAVNKVIAKLDEQERYKSFI 258
L F++++ A+A PK+S +LT+ N+ + ++ E + I
Sbjct: 222 GLVRFAIESEGETGSAIAFFPKDS-ELTEPFNEKLQEMMEDGTMEELI 267

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 223/246 (90%), Positives = 238/246 (96%)

55 Query: 1 MAELKIDVQDLHKSQYQNEVLKGIDAKFYEGDVVCIIGPSGSGKSTFLRTLNLLESITSG 60
M ELKIDVQDLHKSQYQNEVLKGIDAKFYEGDVVCIIGPSGSGKSTFLRTLNLLE+ITSG
Sbjct: 1 MTELKIDVQDLHKSQYQNEVLKGIDAKFYEGDVVCIIGPSGSGKSTFLRTLNLLETITSG 60

60 Query: 61 KVVVDGFELSNPKTDIDKARENIGMVFOHFNLFPHMSVLENIITFAPIELGKESKEAAEKH 120
KV+VDGFELS+PKT+IDKARENIGMVFOHFNLFPHM+VLENI FAP+ELGKESKE A+KH
Sbjct: 61 KVMVDGFELSDPKTNIDKARENIGMVFOHFNLFPHMTVLENIIFAPVELGKESKEVAKKH 120

Query: 121 GMELLEKVGLADKANAKPDSLGGQKQRVAIARSLAMNPDILLFDEPTSAIDPEMVGDLV 180
GM LLEKVGL+DKA+A P SLGGQKQRVAIARSLAMNPDI+LFDEPTSAIDPEMVGDLV
Sbjct: 121 GMALLEKVGLSDKADAFPGSLGGQKQRVAIARSLAMNPDI+LFDEPTSAIDPEMVGDLV 180

-1450-

Query: 181 NVMKDLAEQGMTMLIVTHEMGFARQVANRVIFTDGGRFLEDGTPEQIFDTPQHPRLODFL 240
 NVMKDLAEQGMTMLIVTHEMGFARQVANRVIFTDGG+FLEDGTPE+IFD P+HPRL +FL
 Sbjct: 181 NVMKDLAEQGMTMLIVTHEMGFARQVANRVIFTDGGQFLEDGTPEEIFDHPKHPRLIEFL 240

Query: 241 NKVLNV 246
 +KVLNV
 Sbjct: 241 DKVLNV 246

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1317

A DNA sequence (GBSx1397) was identified in *S.agalactiae* <SEQ ID 4029> which encodes the amino acid sequence <SEQ ID 4030>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2311(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4031> which encodes the amino acid sequence <SEQ ID 4032>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2702(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 45/51 (88%), Positives = 49/51 (95%)

Query: 1 MGDKPISFRDKDGNFVSAADVWNAEKL EELFNTLNPNRKLR LEREKLAKEK 51
 MGDKPISF+DKDGNFVSAADVWNAEKL EELFN LNPNR+LRLEREKL K++
 Sbjct: 11 MGDKPISFKDKDGNFVSAADVWNAEKL EELFNLLNPNRRLRLEREKLKKDE 61

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1318

A DNA sequence (GBSx1398) was identified in *S.agalactiae* <SEQ ID 4033> which encodes the amino acid sequence <SEQ ID 4034>. This protein is predicted to be spo0b-associated GTP-binding protein (obg). Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2967(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1451-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14752 GB:Z99118 GTPase activity [Bacillus subtilis]

Identities = 297/435 (68%), Positives = 345/435 (79%), Gaps = 7/435 (1%)

5
Query: 3 MFLDTAKISVKAGRGDGMVAFRREKYVPNGGPWGGDGGKGGSVIFKVN EGLRTLMDFRY 62
MF+D K+ VK G GG+GMVAFRREKYVP GGP GGDGGKGG V+F+V+EGLRTLMDFRY
Sbjct: 1 MFVDQVKVYVKGDDGGNGMVAFRREKYVPKGGPAGDGGKGGDVVFEVDEGLRTLMDFRY 60

10
Query: 63 NRNFKAKAGEKGMTKGMHGRGAEDLIVSLPPGTTVRDATTGKVITDLVEHDQEFVVARGG 122
++FKA GE GM+K HGR A+D+++ +PPGT V D T +VI DL EH Q V+ARGG
Sbjct: 61 KKHFKAIRGEHGM SKNQHGRNADDMVVKVPPGTTVTD DDTKQVIADLTEHGQRAVIARGG 120

15
Query: 123 RGGGRGNIRFATPRNPAPETAEENGEPGEERELQLELKILADVGLVGFPSVGKSTLLSVVSA 182
RGGGRGN RFATP NPAP+++ENGEPG+ER + LELK+LADVGLVGFPSVGKSTLLSVVS+
Sbjct: 121 RGGGRGNSRFATPANPAPQLSENGEPGKERYIVLELKV LADVGLVGFPSVGKSTLLSVVSS 180

20
Query: 183 AKPKIGAYHFTTIVPNLGMVRTKSGDSFAMADLPGLIEGASQGVGLGTQFLRHIERTRVI 242
AKPKI YHFTT+VFNLMGV T G SF MADLPGLIEGA QGVGLG QFLRHIERTRVI
Sbjct: 181 AKPKIADYHFTTIVPNLGMVETDDGRSFMADLPGLIEGAHQGVGLGHQFLRHIERTRVI 240

25
Query: 243 LHVIDMSASEGRDPYDDYVSINNELETYNLRLMERPQIIIVANKMDMPDSEENLAAFKEKL 302
+HVIDMS EGRDPYDDY++IN EL YNLRL ERPQIIIVANKMDMP++ ENL AFKEKL
Sbjct: 241 VHVIDMSGLEGRDPYDDYLTINQELSEYNLRLTERPQIIIVANKMDMPEAAENLEAFKEKL 300

30
Query: 303 AANYDEFDDMPMIFPISLAHQGLLENLMDATAELLANTEEFLLYDETD MQEDEAYYGFNE 362
DD P +FPIS++ +GL L+ A L NT EF LYDE ++ ++ Y
Sbjct: 301 T-----DDYP-VFPISAVTREGRLRELF EVANQLENTPEFPPLYDEEELTQNRVMTMEN 353

35
Query: 423 GDIVRIGNFEFEFVD 437
GDI+R+ FEFEF+D
Sbjct: 414 GDIIRLLEFEFEFID 428

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4035> which encodes the amino acid sequence <SEQ ID 4036>. Analysis of this protein sequence reveals the following:

40
Possible site: 14
>>> Seems to have no N-terminal signal sequence

45
----- Final Results -----
bacterial cytoplasm --- Certainty=0.2588(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 394/437 (90%), Positives = 421/437 (96%)

50
Query: 1 MSMFLDTAKISVKAGRGDGMVAFRREKYVPNGGPWGGDGGKGGSVIFKVN EGLRTLMDF 60
MSMFLDTAKISV+AGRGDGMVAFRREKYVPNGGPWGGDGGKGGSVIF+V+EGLRTLMDF
Sbjct: 1 MSMFLDTAKISVQAGRGDGMVAFRREKYVPNGGPWGGDGGKGGSVIFRVDEGLRTLMDF 60

55
Query: 61 RYNRNFKAKAGEKGMTKGMHGRGAEDLIVSLPPGTTVRDATTGKVITDLVEHDQEFVVAR 120
RYNR FKAK+GEKGMTKGMHGRGAEDLIV +P GTTVRDA TGKVITDLVEH QE V+A+
Sbjct: 61 RYNRKFKAKSGEKGMTKGMHGRGAEDLIVFPQGTTVRDAETGKVITDLVEHGQEVVIAK 120

60
Query: 121 GGRGGRGNIRFATPRNPAPETAEENGEPGEERELQLELKILADVGLVGFPSVGKSTLLSVV 180
GGRGGRGNIRFATPRNPAPETAEENGEPGEER+L+LELKILADVGLVGFPSVGKSTLLSVV
Sbjct: 121 GGRGGRGNIRFATPRNPAPETAEENGEPGEERQLELELKILADVGLVGFPSVGKSTLLSVV 180

Query: 181 SAAKPKIGAYHFTTIVPNLGMVRTKSGDSFAMADLPGLIEGASQGVGLGTQFLRHIERTR 240
S+AKPKIGAYHFTTIVPNLGMVRTKSGDSFAMADLPGLIEGASQGVGLGTQFLRHIERTR

-1452-

Sbjct: 181 SSAKPKIGAYHFTTIVPNLGMVRTKSGDSFAMADLPGLIEGASQGVGLGTQFLRHIERTR 240

Query: 241 VILHVIDMSASEGRDPYDDYVSINNELETYNLRLMERPQIIIVANKMDMPDSEENLAAFKE 300
 VILHVIDMSASEGRDPY+DYVSINNELETYNLRLMERPQIIIVANKMD+P+++ENL AFK+

5 Sbjct: 241 VILHVIDMSASEGRDPYEDYVSINNELETYNLRLMERPQIIIVANKMDIPEAQENLKAFKK 300

Query: 301 KLAANYDEFDDMPMIFPISSLAHQGLENLMDATAELLANTEEFLLYDETDMQEDEAYYGF 360
 KLA YDEFDD+PMIFPISSLAHQGLENL++ATAELLA T+EFLLYDE+D+ ++EAYYGF

10 Sbjct: 301 KLAAYDEFDDLPMIFPISSLAHQGLENLLEATAELLAKTDEFLLYDESDDLVEEAYYGF 360

Query: 361 NEDERPFEITRDDATWVLYGDKLEKLFVMTNMERDESIMKFARQLRGMGVDEALRERGA 420
 E E+ FEITRDDATWVL G+KLE+LFVMTNMERDESIMKFARQLRGMGVDEALRERGA

Sbjct: 361 AETEKDFEITRDDATWVLSGEKLERLFVMTNMERDESIMKFARQLRGMGVDEALRERGA 420

15 Query: 421 KDGDIVRIGNFEFEFVD 437
 KDGD VRIG FEFEFVD

Sbjct: 421 KDGDPVRIGKFEFEFVD 437

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1319

A DNA sequence (GBSx1399) was identified in *S.galactiae* <SEQ ID 4037> which encodes the amino acid sequence <SEQ ID 4038>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4039> which encodes the amino acid sequence <SEQ ID 4040>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 40 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 30/42 (71%), Positives = 37/42 (87%)

45 Query: 1 MAFGDNQQRKKTGFELTLFVVLMLVLTVGGLVFGAISAIM 42
 +AFG+NG RKKT FEK+T+FVVLMLVLTVGGL+ A+S +M

Sbjct: 1 VAFGENGPRKKTTFEKVTFVVLMLVLTVGGLIASALSVM 42

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1320

A DNA sequence (GBSx1401) was identified in *S.galactiae* <SEQ ID 4041> which encodes the amino acid sequence <SEQ ID 4042>. Analysis of this protein sequence reveals the following:

-1453-

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2484(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAD28348 GB:AF102860 aminopeptidase PepS [Streptococcus hermophilus]
 Identities = 247/413 (59%), Positives = 313/413 (74%)

Query: 1 MVLQDFDNLKKYAQLIISKGLNVQKHTLALTIDVEQVHLARLLTEAAYEKGASEVIVD 60
 MVL +F L+KYA+L+++ G+NVQ GHT+AL+IDVEQ LA LL + AY GA+EVIV

15 Sbjct: 1 MVLPNFKENLEKYAKLLVTNGINVQPGHTVALSIDVEQAEHLAHLVKEAYALGAEEVIVQ 60

Query: 61 YTDDFITRQRLHLSDEVLTNVPQYTVDKSLALLNKKASRLVVKSSNPNAFATVDPKRLS 120
 ++DD I R+R LHA + VP Y + LL KKASRL V+SS+P+AF V P+RLS

20 Sbjct: 61 WSDDTINRERFLHAEMNRIEEVPAYKKAEMEYLLKKASRLGVRSSDPDAFNGVAPERLS 120

Query: 121 ETTRATAIALEEQSRAIQANKVSWNVAAAAGREWAALVFPELKTSDQQVDALWDTIFKLN 180
 +A A + A Q+NKVSU VAAAAG+EWA VFP ++ VD LW+ IFK

25 Sbjct: 121 AHAKAIGAFAKPMQVATQSNKVSUWVAAAAGKEWAKVFPNASSDEEAVDLLWNQIFKTC 180

Query: 181 RIYEDDPAAWDAHEAKLLEKATRLNQEQFDALHYTAPGTDLTGLMPKNHIWEAAGSLNA 240
 R+YE DP+ AW H .+L KA LN+ QF ALHYTAPGTDLTGL+PKNH+WE+AG++NA

30 Sbjct: 181 RVYEKDPVRAWKEHADRLDAKARILNEAQFSALHYTAPGTDLTGLPKNHVWESAGAINA 240

Query: 241 QGETFIANMPTEEIFSAPDYRRADGYVTSTKPLSYAGVIEENMTFTFKDGKIINVTAKEG 300
 QGE+F+ NMPTEE+F+APD+RRA GYV+STKPLSY G IIE + TFKDG+I+++TA++G

35 Sbjct: 241 QGESFLPNMPTEEVFTAPDFRRAYGYVSSTKPLSYNGNIIEGKVTFKDGEIVDITADQG 300

Query: 301 QETVORLIEENDGARSIGLEVALVPHKTPISLSGLIFFNTLFDENASNHLAIGTAYAFNVE 360
 ++ ++ L+ N+GAR+LGE ALVP +PIS SG+ FFNTLFDENASNHLAIG AYA +VE

40 Sbjct: 301 EKVMKNLVFNNGARALGECALVPDSSPISQSGITFFNTLFDENASNHLAIGAAYATSVE 360

Query: 361 GGTEMTSQELDEAGLNRSSTHVDVMIGSEQMDIDGIRADGTAVPIFRNGEWAI 413
 GG +MT +EL AGLNRS HVDF+IGS QM+IDGI DG+ VPIFRNG+W I

45 Sbjct: 361 GGADMTEELKAAGLNRSVDVHVDFIIGSNQMNIDGIIHHDGSRVPIFRNGDWVI 413

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1321

45 A DNA sequence (GBSx1403) was identified in *S.agalactiae* <SEQ ID 4045> which encodes the amino acid sequence <SEQ ID 4046>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -7.91 Transmembrane 661 - 677 (657 - 680)

----- Final Results -----

50 bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55

A related GBS nucleic acid sequence <SEQ ID 8787> which encodes amino acid sequence <SEQ ID 8788> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7

McG: Discrim Score: 6.47

-1454-

GvH: Signal Score (-7.5): 1.01

Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 1 value: -7.91 threshold: 0.0

INTEGRAL Likelihood = -7.91 Transmembrane 658 - 673 (657 - 680)

PERIPHERAL Likelihood = 4.35 555

modified ALOM score: 2.08

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4163(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 647-651

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF09821 GB:AE001885 6-aminohexanoate-cyclic-dimer hydrolase

[Deinococcus radiodurans]

Identities = 150/497 (30%), Positives = 233/497 (46%), Gaps = 32/497 (6%)

Query: 110 LTEETKQKQDGDLANMVRSGQVTSEELVNMAVDIIAKENPSLNAVITTRQEAIEEARK 169

LT Y + D DLA + R G+++E++ A N +LNAV+ + + +AR

Sbjct: 45 LTFAEYDRLDALDLAQLFRRGELSADMDCTAAIHRAQVVNVALNAVYPLYDQGLAQARA 104

Query: 170 L-----KDTNQPFLLGVPLLVKGLGHSIKGGETNNGLIYADGKISTFDSSYVKKYKDLG 222

+ PF GVP LVK G + G G +I +D V++++ G

Sbjct: 105 TDAARARGEQATGPFAGVPFLVKDFGSRLAGVPHTGGTRAYRDQIPEWDELVRWQAAG 164

Query: 223 FIILGQTNFPEYGWRNITDSKLYGLTHNPWDLAHNAGSSSGSAAAIASGMTPIASGSDA 282

+ LG+TN PE+ +T+ +L+G T NPWDL GGSSSGSAA+A+A+G+ P+A D

Sbjct: 165 LLPLGKTNTPFALMGVTEPELHGPTRNPDWDLGRTPGGSSSGSASAVAAGIVPLAGAGDG 224

Query: 283 GGSIRIPSSWTGLVGLKPTRGLV---SNEKPDSTAVHFPLTKSSRDAETLLTYLKKSD 339

GGSIRIP+S GL GLKP+RG V AV LT+S RD+ LL + D

Sbjct: 225 GGSIRIPASCCGLFGLKPSRGRVPCGDGVEPWQGA AVEHVLTRSVRDSAALLDLEQGP 284

Query: 340 QTLVSV-----NDLKSPLIAYTLKSPMGTEVSQDAKNAIMDNVTFLRKQGF 386

+ L I ++ P+G V + A+ L G +

Sbjct: 285 AGAALFLPSPERPYSEEVGREPRLRIGFSTAHPGLGRSVHPECVA AAVQGAARLLES LGHE 344

Query: 387 VTEIDLPIDGRALMRDYSTLAIGMGAFSTIEKDLKKHGFTKEDVDPIWAVHVIYQNSD 446

V E+ LP DG AL + + L G GA +D DV+ +TW + + ++

Sbjct: 345 VEEVALPWDGPALAAQAFMLYFGETGASLAALRDTLGRPARASDVEAVTWLLGQLGRSYS 404

Query: 447 KAEKKKSIMEAQKHMDYRKAMEKLHKQFPPIFLSPTTASLAPLNTDPY----VTEEDKRA 502

A+ A+ + + +AM + H+ + + L+P A+ PL V RA

Sbjct: 405 AAD----FAAARASWNVHARAMGRFHQNYDLLLTPLVAT-PPIQIGELQPRGVQAALLRA 459

Query: 503 IYNMENLSQEERIALFNQWEPMLRRTPFTQIANMTGLPAISIPTYLSESGLPITMLMA 562

M+ R + +L + P+TQ+AN+TG PA+S+P + + GLP+G +A

Sbjct: 460 AQQMDVSGLLRRSGQVDALATDILEKMPYTQLANLTGQPAMSVPLHWTADGLPVGQVQFVA 519

Query: 563 GANYDMVLKIFATFFEK 579

+ VL++ A E+

Sbjct: 520 PLAREDVLLRLAGQLEQ 536

There is also homology to SEQ ID 4048.

SEQ ID 8788 (GBS173) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 5; MW 96.8kDa).

The GBS173-GST fusion product was purified (Figure 116A; see also Figure 201, lane 7) and used to immunise mice (lane 1+2 product; 15µg/mouse). The resulting antiserum was used for Western blot, FACS,

-1455-

and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1322

A DNA sequence (GBSx1404) was identified in *S. agalactiae* <SEQ ID 4049> which encodes the amino acid sequence <SEQ ID 4050>. This protein is predicted to be ribosomal large subunit pseudouridine synthase B (rsuA). Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3674(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06992 GB:AP001518 16S pseudouridylate synthase [Bacillus halodurans]
Identities = 110/236 (46%), Positives = 149/236 (62%), Gaps = 4/236 (1%)

Query: 1  MRIDKFLVECGLSRTQVKLILKKKQISVNGNSETSPKVQVDEYRDEIKYNGTLVSYEKF 60
MR+DKFL  G GSR  VK +LK  + V G      P  V+  + I  G  V Y+  +
Sbjct: 1  MRIDKFLANMGFGSRKDVKKLLKTGAVRVQGPQIKDPSTHVEPESESTITVYGEEVEYKPY 60

Query: 61  VYYMLHKPKGVISATDDPSHKTVLDDLDTARDKAVFPVGRLLDIDTTGLLLLTNNGELAH 120
VY M++KPKGVI AT+D  H+TV+DLL +  R      PVGRLLD DT GLLL+TN+G+  H
Sbjct: 61  VYLMNMKPKGVICATEDLEHETVIDLLGEEERHYEPSVPVGRLLDKDTVGLLLITNDGKFNH 120

Query: 121 KMLSPKKHVDKCYEVKISGIMTEDDILAFDKGIILKD-FTCLPALLEIVEVNQVKQSLV 179
++SPK HV K Y   + G +TE+D+ AF  G++L D +   PA L I+E      +S +
Sbjct: 121 WLMSPKHHVPKTYRALVEGHVTEEDVGAFSHGVLDGYYVTKPATLHILEAG---ARSHI 177

Query: 180 KITIKEGKFHQVKRMVAACGKEVLELKRRLMGNLQLDKQLESGQWRRLTIKEIEKL 235
++ + EGKFHQVKRM  A GK VLEL+R+++GNL LD +L  G++R LT +EI  L
Sbjct: 178 ELILTEGKFHQVKRMFQAVGKRVLELERIKIGNLLLDPELARGEYRELTKEEIALL 233

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4051> which encodes the amino acid sequence <SEQ ID 4052>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0152(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF09821 GB:AE001885 6-aminohexanoate-cyclic-dimer hydrolase
[Deinococcus radiodurans]
Identities = 177/485 (36%), Positives = 259/485 (52%), Gaps = 13/485 (2%)

Query: 5  DATAMAIAVQTGQTTPLLELVTFQAIYKAKKLNPTLNAITSERFEAALEAKQRDPSGL--- 61
DA +A   + G+ +  ++ T AI++A+ +N  LNA+   ++  L +A+  D +
Sbjct: 54 DALDLAQLFRRGELSAEDMCTAAIHRAQVNVNVALNAVVPYLDQGLAQARATDAARARGE 113

Query: 62  ----PFAGVPLFLKDLGQELKGHSSTSGSRLFKKEYQATKTDLFVKRLEALGFIIIGRSNT 117
PFAGVP  +KD G  L G   T G+R +++      D  V+R +A G + LG++NT

```

-1456-

Sbjct: 114 QATGPFAGVPFLVKDFGSRLAGVPHTGGTRAYRDQIPEWDELVRWRQAAGLLPLGKTNT 173

Query: 118 PEFGFKNISDSSLHGPVNLPRDNTRNAGSSSGGAAALVSSGISALATASDGGGSIRIPAS 177
PEF +++ LHGP P D R GGSSGG+A+ V++GI LA A DGGGSIRIPAS

5 Sbjct: 174 PEFALMGVTEPELHGPTRNPWDLGRTPGGSSGGSASAVAAGIVPLAGAGDGGGSIRIPAS 233

Query: 178 FNLGILGLKPSRGRMPVGPFSYRSWQASVHFALTKSVRDTRNLLYYLQMEQMESPFPLAT 237
GL GLKPSRGR+P G G WQGA+V LT+SVRD+ LL Q + L +

10 Sbjct: 234 CCGFLGLKPSRGRVPCGDGVGEPWQGAAVEHVLTSSVRDAAALLDLEQGPDAGAAFLPS 293

Query: 238 LTKDSIYQSLQRP--LTIAFYQRLSDGSPVSLDTAKALRQAVTWLREQGHQQLVELEEFPV 295
+ + + P L I F G V + A++ A L GH++ E+ P

Sbjct: 294 PERPYSEEVGREPGLRIGFSTAHPGRSVHPECVAAVQGAARLLESIGHEVEEV-ALPW 352

15 Query: 296 NMTEVIRHYIIMNSVETAAMFADIEDTFGRPMTKDDMETMTWAIYQSGKDI PAWRYSQVL 355
+ + + + + ET A A + DT GRP D+E +TW + Q G+ A ++

Sbjct: 353 DGPALAQAFMLYFGETGASLAALRDTLGRPARASDVEAVTWLLGQLGRSYSAADFAAAR 412

20 Query: 356 QKWDYTSATMASFHETYDLLLLTFTNTTPAPKHGELVP---DSKLMANLAQAEIFSSEEQF 412
W+ ++ M FH+ YDLLLL TP + GEL P + L+ Q ++ +

Sbjct: 413 ASWNVHARAMGRFHONYDLLLLTFLVLA TPPLQIGELQPRGVQAALLRAAQMDVSGLLRRS 472

Query: 413 NLVETMFGKSLAINPYTALPNLTGQPAISLPTYETKEGLSMGIQLIAAKGREDLLLGIAE 472
V+ + L PYT L NLTGQPA+S+P + T +GL +G+Q +A RED+LL +A

25 Sbjct: 473 GQVDALATDILEKMPYTLQANLTGQPAISVPLHWTADGLPVGQVFAPLAREDVLLRLAG 532

Query: 473 QFEAA 477
Q E A

30 Sbjct: 533 QLEQA 537

An alignment of the GAS and GBS proteins is shown below.

Identities = 151/240 (62%), Positives = 183/240 (75%)

35 Query: 1 MRLDKFLVECGLSRTQVKLILKKKQISVNGNSETSPKVQVDEYRDEIKYNGTLVSYEKF 60
MRLDKFLV G+G+R+QVKL+LKKK I VN ETS K +DEY+D + Y GT + YE F

Sbjct: 2 MRLDKFLVATGVGTRSQVKLILKKKAI FVNQKVETS AKAHIDEYKDLVTYQGTPLVYESF 61

Query: 61 VYYMLHKPKGVISATDDPSHKTVLDLLDKTARDKAVFPVGRLLDITTGLLLLTNNGELAH 120
VYY+L+KP G +SAT D TV++LLD TAR KAVFPVGRLLD DT GLLLLTNNG+LAH

40 Sbjct: 62 VYYLLNKPSGYVSATQDRQATVMELLDDTARQKAVFPVGRLLDKDTRGLLLLTNNGQLAH 121

Query: 121 KMLSPKKHVDKCYEVKISGIMTEDDILAFDKGIILKDFTCPLALLEIVEVNQVKKQSLVK 180
+LSFKKHV K Y K++GIMTE D F +GI LKD CLPA LE++ + ++ SLVK

45 Sbjct: 122 DLLSPKKHVTKEYLAKVAGIMTEADKDYFARGISLKDHCPLPAHLEVLASDLQQQTSLVK 181

Query: 181 ITIKEGKFHQVKRMVAACGKEVLELKRMRMNLQLDKQLESQWRRLTIKEIEKLEKYM 240
ITI+EGKFHQVKRMVAACGKEVL+L+RL MG L+LD L G++RRLT +E++ L Y Q

Sbjct: 182 ITIQEGKFHQVKRMVAACGKEVLDLQRLSMGPLKLDPSLAEGEFRRLTPEELQSLAPYCQ 241

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1323

A DNA sequence (GBSx1405) was identified in *S.agalactiae* <SEQ ID 4053> which encodes the amino acid sequence <SEQ ID 4054>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2811(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1457-

A related GBS nucleic acid sequence <SEQ ID 10007> which encodes amino acid sequence <SEQ ID 10008> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAA57350 GB:J04483 reductase [Leishmania major]
    Identities = 129/277 (46%), Positives = 167/277 (59%), Gaps = 3/277 (1%)

    Query: 26  TLSNTLNIPKIGFGTWQLTEGEEAYKAVTHALKVGYTHIDTAQIYGNEHSVGRAIRDSGL 85
                TLSN + +P+ G G WQ  GE  AV  AL  GY HIDTA IY NE SVG +R SG+
    Sbjct: 10  TLSNGVKMPQFGLGVWQSPAGEVTENAVNWALCAGYRHIDTAAIYKNEESVGAGLRASGV 69

10  Query: 86  ARESIFLTTKIWNDKHDYHLAKASIDESLQKLGVDYIDLLLIHWPNPKALRENDANKAGN 145
                RE +F+TTK+WN + Y  A+ +ES QKLGVDYIDL LIHWP K + + K
    Sbjct: 70  PREDVFITTKLWNTEQGYESTLAAFEESRQKLGVDYIDLYLIHWPGRKDILSKEGKKY-- 127

15  Query: 146 AGTWKAMEEAYKEGKVKVKAIGVSNFMKHHLEALFETAEIKPMVNQIILAPGCAQEDLVRFC 205
                +W+A E+ YKE KV+AIGVSNF HHLE + + PMVNQ+ L P  Q DL FC
    Sbjct: 128 LDSWRAFEQLYKEKKVRAIGVSNFHHLEEDVLAMCTVTPMVNQVELHPLNNQADLRAFC 187

    Query: 206 KGNDILLEAYSFPFGTGAIFENESIKATAEKYKGSVAQVALRWSLDNGFLPLPKSATPKNI 265
                I +EA+SP G G + N + AI  KY K+ AQV LRW++ + +PKS + I
    Sbjct: 188 DAKQIKVEAWSPLGQGLLSNPILSAIGAKYNKTAQVILRWNIQKNLITIPKSVHRERI 247

    Query: 266 EANLDIFDFQLNEDDIATLIQLDSGIK-PKDPDNVSF 301
                E N DIFDF+L +D+ ++ L++ + DPD F
25  Sbjct: 248 EENADIFDFELGAEDVMSIDALNTNSRYGPDPEAQF 284

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 779> which encodes the amino acid sequence <SEQ ID 780>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 27
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.0980(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 155/282 (54%), Positives = 204/282 (71%), Gaps = 2/282 (0%)

40  Query: 20  IVMETYTLNIPKIGFGTWQLTEGEEAYKAVTHALKVGYTHIDTAQIYGNEHSVGRA 79
                +++ T +++ IP +GFGT+Q +GEEAY++ A+K GY HIDTA IY NE SVGRA
    Sbjct: 1  VMVTTVKMTSGYEIPVLGFGTYQAADGEEAYQSTLAAIKAGYRHIDTAAIYKNEESVGRA 60

    Query: 80  IRDSGLARESIFLTTKIWNDKHDYHLAKASIDESLQKLGVDYIDLLLIHWPNPKALREND 139
                I+DSG+ RE +F+TTK+WND H Y AK ++ SL +LG+DY+DL LIHWPNPKALR +
45  Sbjct: 61 IKDSGVLRDLFITTKLWNDASHYEGAKDALAASLDRLGLDYVDLYLIHWPNPKALR--N 118

    Query: 140 AWKAGNAGTWKAMEEAYKEGKVKVKAIGVSNFMKHHLEALFETAEIKPMVNQIILAPGCAQE 199
                WK NA W+ MEEA + G +K+IGVSNFM HHLEAL ETA+I P +NQI LAPGC Q+
50  Sbjct: 119 TWKEANAQAWQYMEEAVEAGLIKSIGVSNFMVHHLEALQETAKITPAINQIRLAPGCYQK 178

    Query: 200 DLVRFC KGNDILLEAYSFPFGTGAIFENESIKATAEKYKGSVAQVALRWSLDNGFLPLPKS 259
                ++V +CK N+ILLEA+SP G G IF+NE+++ +A KY K+VAQVAL WSL GF+PLPKS
    Sbjct: 179 EVVDYCKANEILLEAWSPLGQGEIFDNETMQQLANKYDKTVAQVALAWSLAEGFIPLPKS 238

55  Query: 260 ATPKNIEANLDIFDFQLNEDDIATLIQLDSGIKPKDPDNVSF 301
                + I+ N+ IFD L ++D T+ L +PD SF
    Sbjct: 239 VHDERIKENMAIFDVSLTQEDDKKTIRYLSGMSAIPNPDTSF 280

```

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1458-

Example 1324

A DNA sequence (GBSx1406) was identified in *S.agalactiae* <SEQ ID 4055> which encodes the amino acid sequence <SEQ ID 4056>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0633(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10009> which encodes amino acid sequence <SEQ ID 10010> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12612 GB:Z99108 similar to NAD(P)H-flavin oxidoreductase

[Bacillus subtilis]

Identities = 106/223 (47%), Positives = 150/223 (66%), Gaps = 8/223 (3%)

Query: 29 DIKKQVRRAFDFRMAIRVYN--NNDIPKEDMEYILDTAWLSPSSVGLGWRFLVLDRTQIA 87

D+K Q+ A++FR A + ++ N + D E+IL+T LSPSS+GLE W+F+V+

Sbjct: 3 DLKTQILDAYNFRHATKEFDPNKKVSDSDFEFILETGRLSPSSLGLEPWKFVVVQNP--- 59

Query: 88 KFRDKLKEVAWGAQYQLDTASHFVLLAE--KGAYYNADSMINSLIRRLGLGDPAALESRI 145

+FR+KL+E WGAQ QL TASHFVL+LA K YNAD + L E +

Sbjct: 60 EFREKLREYTWGAQKQLPTASHFVLILARTAKDIKYNADYIKRHLKEVKQMPQDVYEGYL 119

Query: 146 PLYKSFQENDMKI-DSERSLWDWTAKQTYIALGNMNTAAAMIGVDSCPIEGFDYKVNMI 204

+ FQ+ND+ + +S+R+L+DW +KQTYIALGNMNTAAA IGVDSCPIEGF Y+ ++ I

Sbjct: 120 SKTEEFQKNDLHLLSDRTLFDWASKQTYIALGNMNTAAAIQIGVDSCPIEGFYDHIHRI 179

Query: 205 LSKEGLIDDKKEAISCMVSFGYRLREP KHSRARKERQEVITWV 247

L +EGL+++ IS MV+FGYR+R+P+ + R ++V+ WV

Sbjct: 180 LEEGLLENGSFDISVMVAFGYRVRDPR-PKTRSAVEDVVKWV 221

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4057> which encodes the amino acid sequence <SEQ ID 4058>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1705(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 126/222 (56%), Positives = 174/222 (77%), Gaps = 4/222 (1%)

Query: 28 EDIKKQVRRAFDFRMAIRVYNNNDIPKEDMEYILDTAWLSPSSVGLGWRFLVLDRTQIA 87

+ I Q+++A FR A+RVY I ED+ ILD AWLSPSS+GLEGWRFLVLD + I

Sbjct: 3 QTIIHQIQALHFRITAVRVYKEEKISDEDLALILDAWLSPSSIGLEGWRFLVLDNKPI- 61

Query: 88 KFRDKLKEVAWGAQYQLDTASHFVLLAEKGAYYNADSMINSLIRRLGLGDPAALESRIPL 147

++++K AWGAQYQL+TASHF+LL+AEK A Y++ ++ NSL+RRG+ + L SR+ L

Sbjct: 62 --KEEIKPFAWGAQYQLE+TASHFILLIAEKHARYDSPAIKNSLLRRGIKEGDGLNSRLKL 119

Query: 148 YKSFQENDMKI-DSERSLWDWTAKQTYIALGNMNTAAAMIGVDSCPIEGFDYKVNMI 206

Y+SFQ+ DM + D+ R+L+DWTAKQTYIALGNMM AA++G+D+CPIEGF Y+KVN+IL+

Sbjct: 120 YESFQKEDMDMADNPRLFDWTAKQTYIALGNMNTAALLGIDTCPIEGFHYDKVNHILA 179

-1459-

Query: 207 KEGLIDDKKEAISCMVSFGYRLREP KHSRARKERQEVITWVE 248
 K +ID +KE I+ M+S GYRLR+PKH++ RK ++EVI+ V+
 Sbjct: 180 KHNVIDLEKEGIASMLSLGYRLRDPKHAQVRKPKEEVISVVK 221

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1325

A DNA sequence (GBSx1407) was identified in *S.agalactiae* <SEQ ID 4059> which encodes the amino acid sequence <SEQ ID 4060>. This protein is predicted to be lactoylglutathione lyase (gloA). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1656(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAC21986 GB:U32717 lactoylglutathione lyase (gloA) [Haemophilus influenzae Rd]
 Identities = 59/131 (45%), Positives = 86/131 (65%), Gaps = 2/131 (1%)

Query: 1 MPFLHTCIRVKDLDAISAFYQEALGFKEVRRNDFPENQFTLVYMALEDDPSY-ELELTYN 59
 M LHT +RV DLD SI FYQ+ LG + +R ++ PE ++TL ++ ED S E+ELTYN
 25 Sbjct: 1 MQILHTMLRVGDLDRSIKFYQDVLGMRLRLTSENPEYKYTLAFLGYEDGESAAEIELTYN 60
 Query: 60 YDHEAYDLGNGYGHIAVGVDLETTYDAHQKAGYSVTKISG-LPGKPNMFYFIQDPDGYK 118
 + + Y+ G YGHIA+GVDD+ T +A + +G +VT+ +G + G + F++DPDGYK
 30 Sbjct: 61 WGVDKYEHGTAYGHIAIGVDDIYATCEAVRASGGNVTREAGPVKGGSTVIAFVEDPDGYK 120
 Query: 119 IEVIRLSQFKA 129
 IE I K+
 Sbjct: 121 IEFIENKSTKS 131

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4061> which encodes the amino acid sequence <SEQ ID 4062>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1382(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 80/125 (64%), Positives = 93/125 (74%), Gaps = 1/125 (0%)

Query: 1 MPFLHTCIRVKDLDAISAFYQEALGFKEVRRNDFPENQFTLVYMALEDDPSY-ELELTYN 60
 M LHTCIRVKDLD S+AFY A FKE R DFP++QFTLVY+ALE + SYELELTYN
 50 Sbjct: 1 MKALHTCIRVKDLDQSVAFYTSAPFPFKENYRKDFPDSQFTLVYLALEGE-SYELELTYN 59
 Query: 61 DHEAYDLGNGYGHIAVGVDLETTYDAHQKAGYSVTKISGLPGKPNMFYFIQDPDGYKIE 120
 H YDLGNGYGHIA+G + E + H++AG+ VT I L K +YFIQDPDGYKIE
 Sbjct: 60 GHGDYDLGNGYGHIALGSEHFADHKKHRQAGFPVTDIKELADKSARYYFIQDPDGYKIE 119
 55 Query: 121 VIRLS 125
 VI L+
 Sbjct: 120 VIDLN 124

-1460-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1326

- 5 A DNA sequence (GBSx1408) was identified in *S.agalactiae* <SEQ ID 4063> which encodes the amino acid sequence <SEQ ID 4064>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have an uncleavable N-term signal seq

10 INTEGRAL Likelihood = -9.02 Transmembrane 241 - 257 (229 - 262)
INTEGRAL Likelihood = -4.94 Transmembrane 270 - 286 (264 - 287)

----- Final Results -----

15 bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12688 GB:Z99108 stress response protein [Bacillus subtilis]

Identities = 139/304 (45%), Positives = 200/304 (65%), Gaps = 3/304 (0%)

20 Query: 3 LLSVIVPCYNEQETVSTFLTEIKKVESEMARYTHFEYIFVNDGSTDRTELELLKKAQKQFD 62
L+S+I+P YNE V +KK E + Y +E F+NDGS D TL+ +K A
Sbjct: 5 LISIIIPSYNEGYNVKLIHESLKK-EFKNIHYD-YEIFFINDGSVDDTLQQIKDLAATCS 62

25 Query: 63 NVHYLSFSRHFQKDAALLAGLEHTTGDFITVMDVDLQDPPTLLPEMYLKLQEGYDIVATR 122
V Y+SFSR+FGK+AA+LAG EH G+ + VMD DLQ P LL E +EGYD V +
Sbjct: 63 RVKYISFSRNFQKEAAILAGFEHVQGEAVIVMDADLQHPTYLLKEFIKGYEEGYDQVIAQ 122

30 Query: 123 RKDRKGEPILIRSLFAKLFYKLINQVSDTKMVDGARDFRLMTKQVVDLSILELNEVNRFSKG 182
R +RKG+ +RSL + ++YK IN+ + + DG DFRL+++Q V+++L+L+E NRFSKG
Sbjct: 123 R-NRKGDSFVRSLLSSMYKFKINAVEVDLRDGVGDFRLLSRQAVNALLKLSEGNRFSKG 181

35 Query: 183 IFSWIGYDVAYISYENRERIAGKTSWSFFNLLKYSLDGFINFSEIPLAIATWIGTLSSVL 242
+F WIG+D + YEN ER G + WSF +L Y +DG ++F+ PL + + G +L
Sbjct: 182 LFCWIGFDQKIVFYENVERKNGTSKWSFSSLFNYGMDGVVSFNHKPLRLCFYTGIFILL 241

40 Query: 243 SLLAIIFIIIRKLLFGDPVSGWASTVTIVLFMGGIQLLSLGIIGKYISKIFLETKKRPVY 302
S++ II ++ L G V G+ + ++ VLF+GG+QLLSLGIIG+YI +I+ ETKKRP Y
Sbjct: 242 SIIYIIATFVKILTNGISVPGYFTIISAVLFLGGVQLLSLGIIGEYIGRIYYETKKRPHY 301

Query: 303 IVKE 306
++KE
Sbjct: 302 LIKE 305

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4065> which encodes the amino acid sequence <SEQ ID 4066>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have an uncleavable N-term signal seq

50 INTEGRAL Likelihood = -9.55 Transmembrane 256 - 272 (251 - 282)
INTEGRAL Likelihood = -5.31 Transmembrane 290 - 306 (284 - 307)

----- Final Results -----

55 bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9113> which encodes the amino acid sequence <SEQ ID 9114>. Analysis of this protein sequence reveals the following:

-1461-

Possible cleavage site: 36
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty= 0.482(Affirmative) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 207/307 (67%), Positives = 258/307 (83%)

Query: 1 MALLSVIVPCYNEQETVSTFLTEIKKVESEMARYTHFEYIFVNDGSTDRTELKKAQ 60
 M LLS+IVPC+NE+ + + E+ ++E+ M FEYIF++DGS D TL +L++ A +
 15 Sbjct: 21 MTLLSIIVPCFNEEANILPYFEEMHQLETSMTNQLAFEYIFIDDGSKDNTLGILRELAAR 80

Query: 61 FDNVHYLSFSRHFQKDAALLAGLEHTTGDFITVMDVDLQDPPPTLLPEMYLKLQEGYDIVA 120
 F NVHYLSFSRHFQK+A LLAGL+ G++ITVMDVDLQDPP LLP MY KL+EGYDIV
 20 Sbjct: 81 FPNVHYLSFSRHFQKEAGLLAGLKEAKGNYITVMDVDLQDPELLPIMYAKLKEGYDIVG 140

Query: 121 TRRKDRKGEPLIRSLFAKLFYKLINQVSDTKMVDGARDFRLMTKQVVDLSILELNEVNRFS 180
 TRR++R+GEPLIRS+ + LFY LI +SDT+MV+G RD+RLMT+QVVDLSILEL EVNRFS
 25 Sbjct: 141 TRRQNRQGEPLIRSMCSNLFYGLIKHLSDETMVNGVRDYRLMTRQVVDLSILELGEVNRFS 200

Query: 181 KGIFSWIGYDVAYISYENRERIAGKTSWSFFNLLKYSLDGFINFSEIPLAIATWIGTLSS 240
 KGIFSW+GY + Y+S+EN++R GK+ W F+ LL+YSLDGFINFSE+PL IATW GT S
 30 Sbjct: 201 KGIFSWVGYRITYLSFENQKRKYGKSRWHFWELLRYSLDGFINFSEMPITATWTGTFSF 260

Query: 241 VLSLLAIIFIIIRKLLFGDPVSGWASTVTIVLFMGGIQLLSLGIIGKYISKIFLETKKRP 300
 ++S+ AI+PIIIRK+LFGDPVSGWASTV+I+LFMGGIQL +GIIGKYISKIFLETKKRP
 35 Sbjct: 261 LISIFAILFIIIRKILFGDPVSGWASTVSIILFMGGIQLFCMGIIGKYISKIFLETKKRP 320

Query: 301 VYIVKEE 307
 +YI+KE+
 40 Sbjct: 321 LYIIKEK 327

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1327

40 A DNA sequence (GBSx1409) was identified in *S.agalactiae* <SEQ ID 4067> which encodes the amino acid sequence <SEQ ID 4068>. This protein is predicted to be d-serine/d-alanine/glycine transporter (cycA). Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -2.44 Transmembrane 50 - 66 (50 - 66)
 INTEGRAL Likelihood = -1.49 Transmembrane 27 - 43 (27 - 43)

----- Final Results -----

50 bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA83253 GB:Z31377 potential amino acid permease
 [Lactobacillus delbrueckii]

55 Identities = 34/55 (61%), Positives = 44/55 (79%)

Query: 7 DHTQKSENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSISLTGPSIVLVYAITG 61
 D + ++ +G +R L NRHVQ+IAI GTIGTGLFLGAG +IS TGPS++ +YAI G
 60 Sbjct: 5 DRSIENTDGTIRSLSNRHHVQMIAGGTIGTGLFLGAGTTISATGPSVIFYAIMG 59

-1462-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4069> which encodes the amino acid sequence <SEQ ID 4070>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

5	INTEGRAL	Likelihood = -11.15	Transmembrane	170 - 186 (161 - 190)
	INTEGRAL	Likelihood = -8.44	Transmembrane	256 - 272 (252 - 274)
	INTEGRAL	Likelihood = -8.33	Transmembrane	352 - 368 (347 - 375)
	INTEGRAL	Likelihood = -7.54	Transmembrane	139 - 155 (133 - 160)
	INTEGRAL	Likelihood = -5.73	Transmembrane	420 - 436 (417 - 440)
10	INTEGRAL	Likelihood = -3.88	Transmembrane	56 - 72 (54 - 75)
	INTEGRAL	Likelihood = -3.40	Transmembrane	283 - 299 (282 - 300)
	INTEGRAL	Likelihood = -3.29	Transmembrane	440 - 456 (439 - 458)
	INTEGRAL	Likelihood = -1.49	Transmembrane	31 - 47 (31 - 47)
15	INTEGRAL	Likelihood = -1.33	Transmembrane	109 - 125 (109 - 127)

----- Final Results -----

bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14651 GB:Z99117 amino acid permease [Bacillus subtilis]
Identities = 210/454 (46%), Positives = 296/454 (64%), Gaps = 11/454 (2%)

25	Query: 12	DNNELENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSIALTGPSIIFVYMITGAFMFM 71
	DN	+ + RGL+NRH+QL+AI G IGTGLFLG+G+SI GPSI+F Y+ITG F F +
	Sbjct: 8	DNFGQQQKLSRGLKNRHIQLMAIGGAIGTGLFLGSGKSIHFAGPSILFAYLITGVFCFFI 67
30	Query: 72	MRAIGEMLYDDPDQHTFINFISKYIGPGWGYFSGLSYWSLIFIGMAEITAVGAYVQFWF 131
		+R++GE+L + H+F++F+ Y+G + +G +YW I + MA++TAVG Y Q+W
	Sbjct: 68	IRSLGELLLSNAGYHSFVDFVRDYLGNMAAFITGWTYWFCWISLAMADLTAVGIYQTQWL 127
35	Query: 132	PSWPAWLIQLVFLVLLSSINLIAVRVFGETEWFAMIKILAILALIATAIFMVLTFGETH 191
		P P WL L+ L++L +NL V++FGE EFWFA+IK++AILALI T I ++ GF
	Sbjct: 128	PDVPPQWLPGLLALIILLIMNLATVKLFGELEFWFALIKVIAILALIVTGILLIAKGFSA 187
40	Query: 192	TGHASLSNIFDHFMSFPNGKLFKFMFQMVFFAYQAIEFVGITTTSETANPRKVLPAKIQE 251
		+G ASL+N++ H MFPNG F ++FQMV FA+ IE VG+T ET NP+KV+PKAI +
	Sbjct: 188	SGPASLNNLWSHGGMFPNGWHGFILSFQMVVFAFVGIELVGLTAGETENPQKVIPKAINQ 247
45	Query: 252	IPTRIVIFYVGALVSIMAIVPWHQLPVDSPFVMVFKLIGIKWAAALINFVVLTSAAASAL 311
		IP RI++FYVGAL IM I PW+ L +ESPFV VF +GI AA+LINFVVLTSAAASA
	Sbjct: 248	IPVRILLFYVGALFVIMCIYPWNVLNPNESPFVQVFSAVGIVVAASLINFVVLTSAAASAA 307
50	Query: 312	NSTLYSTGRHLYQIANE--TPNALTNRLKINTLSRQGVPSRAIASAVVVGISALINILP 369
		NS L+ST R +Y +A + P L L+ VPS A+ S++ + I +N L
	Sbjct: 308	NSALFSTSRMVYSLAKDHHAPGLL-----KKLTSSNVPSNALFFSSIAILIGVSLNYLM 361
55	Query: 370	GVADAFSLITASSSGVYIAIYALTMIAHWKYRQSK--DFMADGYLMPKYKVTTPLTLAFL 427
		F+LIT+ S+ +I I+ +T+I H KYR+++ + A+ + MP Y ++ LTLAF
	Sbjct: 362	-PEQVFTLITSVSTICFIFIWGITVICHLYKRYKTRQHEAKANKFKMPFYPLSNYLTFLAFL 420
	Query: 428	AFVFISLFLQESTYIGAIGATIWIIFGIYSNVK 461
		AF+ + L L T I +W ++ I V+
	Sbjct: 421	AFILVILALANDTRIALFVTPVWFVLLIILYKVQ 454

An alignment of the GAS and GBS proteins is shown below.

Identities = 48/62 (77%), Positives = 51/62 (81%)

60	Query: 1	MSKNNNDHTQKSENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSISLTGPSIVLVYAITGA 62
	MS	+ ENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSI+LTGPSI+ VY ITGA
	Sbjct: 5	MSIKEQTDNNELENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSIALTGPSIIFVYMITGA 66

-1463-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1328

A DNA sequence (GBSx1411) was identified in *S.agalactiae* <SEQ ID 4071> which encodes the amino acid sequence <SEQ ID 4072>. This protein is predicted to be alkylphosphonate uptake protein (phnA). Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0965(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77069 GB:AE000483 orf, hypothetical protein [Escherichia coli K12]
Identities = 79/110 (71%), Positives = 91/110 (81%), Gaps = 1/110 (0%)

Query: 1 MSLPNCPCNCSEYVYEDGILLVCPECAYEWNPEE-IEEEVGLIVLDSNGTRLSDGDTVTV 59
MSLP+CPKCNCSEY YED + +CPECAYEWN E +E LIV D+NG L+DGD+VT+
Sbjct: 1 MSLPHCPKCNCSEYTYEDNGMYICPECAYEWNDAEPAQESDELIVKDANGNLLADGDSVTI 60

Query: 60 IKDLKVKGAPKDIKQGTRVKNIRLVDGDHNIDCKIDGFGAMKLKSEFVKK 109
IKDLKVKG+ +K GT+VKNIRLV+GDHNIDCKIDGFG MKLKSEFVKK
Sbjct: 61 IKDLKVKGSSSMLKIGTKVKNIRLVEGDHNIDCKIDGFGPMKLKSEFVKK 110

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4073> which encodes the amino acid sequence <SEQ ID 4074>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3428(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 73/85 (85%), Positives = 79/85 (92%), Gaps = 1/85 (1%)

Query: 26 CAYEWNP-EEIEEEVGLIVLDSNGTRLSDGDTVTVIKDLKVKGAPKDIKQGTRVKNIRLV 84
CA+EW P EE EE GL+VLDNG RLSDGDT+TV+KDLKVKGAPKD+KQGTRVKNIRLV
Sbjct: 2 CAFEWTPGEEATEEEGLVVLDSNGVRLSDGDTITVVKDLKVKGAPKDLKQGTRVKNIRLV 61

Query: 85 DGDHNIDCKIDGFGAMKLKSEFVKK 109
+GDHNIDCKIDGFGAMKLKSEFVKK
Sbjct: 62 EGDHNIDCKIDGFGAMKLKSEFVKK 86

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1329

A DNA sequence (GBSx1412) was identified in *S.agalactiae* <SEQ ID 4075> which encodes the amino acid sequence <SEQ ID 4076>. Analysis of this protein sequence reveals the following:

Possible site: 22

-1464-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3665(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 500.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1330

A DNA sequence (GBSx1414) was identified in *S.agalactiae* <SEQ ID 4077> which encodes the amino acid sequence <SEQ ID 4078>. Analysis of this protein sequence reveals the following:

15 Possible site: 13
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.11 Transmembrane 558 - 574 (558 - 574)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:CAB11971 GB:Z99105 L-glutamine-D-fructose-6-phosphate
 amidotransferase [Bacillus subtilis]
 Identities = 355/604 (58%), Positives = 445/604 (72%), Gaps = 4/604 (0%)

30 Query: 1 MCGIVGVGNTNATDILIQGLEKLEYRGYDSAGIFVVGDNKSQLVKSVGRIAEIQAKVGD 60
 MCGIVG +G +A +IL++GLEKLEYRGYDSAGI V + + K GRIA+++ V
 Sbjct: 1 MCGIVGYIGQLDAKEILLKGLEKLEYRGYDSAGIAVANEQGIHVFEKEGRIADLREVVDA 60

35 Query: 61 SVSGTTGIGHTRWATHGKPTGNAHPHTSGSGRFLVHNGVIENYLQIKETYLTKHNLKG 120
 +V GIGHTRWATHG+P+ NAHPH S GRF LVHNGVIENY+Q+K+ YL LK
 Sbjct: 61 NVEAKAGIGHTRWATHGEPSTYLNAPHQALGRFTLVHNGVIENYVQLKQEYLQDVELKS 120

40 Query: 121 ETDTEIAIHLVEHFVEEDNLSVLEAFKKALHIIEGSYAFALIDSQDADTIYVAKNKSPLL 180
 +TDTE+ + ++E FV L EAF+K L +++GSYA AL D+ + +TI+VAKNKSPLL
 Sbjct: 121 DTDTEVVVQVIEQFVN--GLETEEAFRKTLTLLKGSYAIALFDNDNRETIFVAKNKSPLL 179

45 Query: 181 IGLNGNYMVCSDAMAMIRETSEYMEIHDKELVIVKKDSVEVQDYDGNVIERGSGSYTAELD 240
 +GLG+ +N+V SDAMAM++ T+EY+E+ DKE+VIV D V +++ DG+VI R SY AELD
 Sbjct: 180 VGLGDTFNVVASDAMAMLQVTNEYVELMDKEMVIVTDDQVVIKNLDGDVITRASVIAELD 239

50 Query: 241 LSDIGKGTYPFMYLKEIDEQPTVMRKLITYANESGDMNVSDIIKSVQEADRLYILAAG 300
 SDI KGTYP YMLKE DEQP VMRK+I TY +E+G ++V DI +V EADR+YI+ G
 Sbjct: 240 ASDIEKGTYPHYMLKETDEQPVMRKIIQTYQDENGKLSVPGDIAAAVAEADRIYIIGCG 299

55 Query: 301 TSYHAGFAAKTMIEKLTDTFVELGVSSEWGYNMPLLSKKPMFILLSQSGETADSRQVLVK 360
 TSYHAG K IE + PVE+ V+SE+ YNMPLLSKKP+FI LSQSGETADSR VLV+
 Sbjct: 300 TSYHAGLVGKQYIEMWANVPVEVHVASEFSYNMPLLSKKPLFIFLSQSGETADSRVAVLVQ 359

60 Query: 361 ANEMGIPSLTITNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQVATLAFLAKAVGEA 420
 +G +LTITNVPGSTLSREA YT+L+HAGPEIAVASTKAYTAQ+A LA LA +
 Sbjct: 360 VKALGHKALTITNVPGSTLSREADYTLLHAGPEIAVASTKAYTAQIAVLAVLASVAADK 419

60 Query: 421 NGKAEAKDFDLVHELISIVAQSIEATLSEKDVISEKVEQLLISTRNAFYIGRGNDYYVTME 480
 NG FDLV EL I A ++EA +KD + + L +RNAF+ICRG DY+V +E
 Sbjct: 420 NGINIG--FDLVKELGIAANAMEALCDQKDEMEMIAREYLTVSRNAFFIGRGLDYFVCVE 477

-1465-

Query: 481 AALKLKEISYIQTEGFAAGELKHGTISLIEDNTPVIALISADSTIAAHTRGNIQEVVSRG 540
 ALKLKEISYIQ EGFA GELKHGTI+LIE TPV AL + + + RGN++EV +RG
 Sbjct: 478 GALKLKEISYIQAEFGAGGELKHGTIALIEQGTVPFALATQEH-VNLSIRGNVKEVAARG 536

5 Query: 541 ANALIIVEEGLEREGDDIIVNKVHPFLSAISMVIPTQLIAYYASLQRGLDVKPRNLAKA 600
 AN II +GL+ .D ++ +V+P L+ + V+P QLIAYYA+L RG DVDKPRNLAK+
 Sbjct: 537 ANTCTISLKGGLDDADDRFVLPEVNPALAPLVSVVPLQLIAYYAALHRGCDVDKPRNLAKS 596

10 Query: 601 VTVE 604
 VTVE
 Sbjct: 597 VTVE 600

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4079> which encodes the amino acid sequence <SEQ ID 4080>. Analysis of this protein sequence reveals the following:

15 Possible site: 39
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.06 Transmembrane 558 - 574 (558 - 574)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:CAB11971 GB:Z99105 L-glutamine-D-fructose-6-phosphate
 amidotransferase [Bacillus subtilis]
 Identities = 353/604 (58%), Positives = 445/604 (73%), Gaps = 4/604 (0%)

30 Query: 1 MCGIVGVGNRNATDILMQGLEKLEYRGYDSAGIFVANANQTNLIKSVGRIADLRKIGI 60
 MCGIVG +G +A +IL++GLEKLEYRGYDSAGI VAN ++ K GRIADLR +
 Sbjct: 1 MCGIVGYIGQLDAKEILLKGLEKLEYRGYDSAGI AVANEQGIHVFKKGRIADLRVVD 60

35 Query: 61 DVAGSTGIGHTRWATHGQSTEDNAHPHTSQTGRFVLVHNGVIENYLHIKTEFLAGHDFKG 120
 +V GIGHTRWATHG+ + NAHPH S GRF LVHNGVIENY+ +K E+L + K
 Sbjct: 61 NVEAKAGIGHTRWATHGEPSTYLAHPHQSALGRFTLVHNGVIENYVQLKQEYLQDVELKS 120

40 Query: 121 QTDTEIAVHLIGKFVEEDKLSVLEAFKKSLSIIEGSAFALMDSQATDTIYVAKNKSPLL 180
 TDTE+ V +I +FV L EAF+K+L++++GSYA AL D+ +TI+VAKNKSPLL
 Sbjct: 121 DTDTEVVVQVIEQFVNGG-LETEEAFRKTLLTLKGSYAIALFDNDNRETIFVAKNKSPLL 179

45 Query: 181 IGLGEGYNMVCSDAMAMIRETSEFMEIHDKELVILTKDKVTVDYDGKELIRDSYTAELD 240
 +GLG+ +N+V SDAMAM++ T+E++E+ DKE+VI+T D+V + + DG + R SY AELD
 Sbjct: 180 VGLGDTFNVVASDAMAMQVTNEYVELMDKEMVIVTDDQVVIKNLDGDVITRASIAELD 239

50 Query: 241 LSDIGKGTYPFYMLKEIDEQPTVMRQLISTYADETGNVQVDPAIITSIQEADRILYILAAG 300
 SDI KGTYP YMLKE DEQP VMR++I TY DE G + V I ++ EADR+YI+ G
 Sbjct: 240 ASDIEKGTYPHYMLKETDEQPVVMRKIIQTYQDENGKLSVPGDIAAAVAEADRIYIIGCG 299

55 Query: 301 TSYHAGFATKNMLEQLTDTFVELGVASEWGYHMLLSKKPMFILLSQSGETADSRQVLVK 360
 TSYHAG K +E + PVE+ VASE+ Y+MPLLSKKP+FI LSQSGETADSR VLV+
 Sbjct: 300 TSYHAGLVGKQYIEMWANVPVEVHVASEFSYNMPLLSKKPLFIFLSQSGETADSRVLVQ 359

60 Query: 361 ANAMGIPSLTVTNVPGSTLSREATYTMILHAGPEIAVASTKAYTAQIAALAFIAKAVGEA 420
 A+G +LT+TNVPGSTLSREA YT+L+HAGPEIAVASTKAYTAQIA LA LA +
 Sbjct: 360 VKALGHKALTITNVPGSTLSREADYTLLHAGPEIAVASTKAYTAQIAVLAVLASVAADK 419

65 Query: 421 NGKQEAELDFNLVHELSLVAQSIEATLSEKDLVAEKVQALLATTFNAFYIGRGNDYYVAME 480
 NG + F+LV EL + A ++EA +KD + + L +RNAF+IGRG DY+V +E
 Sbjct: 420 NGIN--IGFDLVKELGIAANAMEALCDQKDEMEMIAREYLTVSRNAFFIGRGLDYFVCVE 477

Query: 481 AALKLKEISYIQCEGFAAGELKHGTISLIEEDTPVIALISSSQLVASHTRGNIQEVARG 540
 ALKLKEISYIQ EGFA GELKHGTI+LIE+ TPV AL + + S RGN++EVAARG
 Sbjct: 478 GALKLKEISYIQAEFGAGGELKHGTIALIEQGTVPFALATQEHVNLS-IRGNVKEVAARG 536

Query: 541 AHVLTVVEEGLDREGDDIIVNKVHPFLAPIAMVIPTQLIAYYASLQRGLDVKPRNLAKA 600

-1466-

A+ + +GLD D ++ +V+P LAP+ V+P QLIAYYA+L RG DVDKPRNLAK+
 Sbjct: 537 ANTCTIISLKLDDADDRFVLPEVNPALAPLVSVVPLQLIAYYAALHRGCDVDKPRNLAKS 596

Query: 601 VTVE 604
 VTVE
 Sbjct: 597 VTVE 600

An alignment of the GAS and GBS proteins is shown below.

Identities = 500/604 (82%), Positives = 552/604 (90%)

Query: 1 MCGIVGVVGNATDILIQGLEKLEYRGYDSAGIFVVGDNKSQLVKSVGRIAEIQAKVGD 60
 MCGIVGVVGN NATDIL+QGLEKLEYRGYDSAGIFV N++ L+KSVGRIA+++AK+G
 Sbjct: 1 MCGIVGVVGNRNATDILMQGLEKLEYRGYDSAGIFVANANQTNLIKSVGRIADLRKIGI 60

Query: 61 SVSGTTGIGHTRWATHGKPTTEGNAHPHTSGSGRFVLVHNGVIENYLQIKETYLTKNLKG 120
 V+G+TGIGHTRWATHG+ TE NAHPHTS +GRFVLVHNGVIENYL IK +L +H+ KG
 Sbjct: 61 DVAGSTGIGHTRWATHGQSTEDNAHPHTSQTGRFVLVHNGVIENYLHIKTEFLAGHDFKG 120

Query: 121 ETDTEIAIHLVEHFVEEDNLSVLEAFKKALHIEGSAFALIDSQDADTIYVAKNKSPLL 180
 +TDTEIA+HL+ FVEED LSVLEAFKK+L IIEGSAFAL+DSQ DTIYVAKNKSPLL
 Sbjct: 121 QTDTEIAVHLIGKFVEEDKLSVLEAFKKSLSIEGSAFALMDSQATDTIYVAKNKSPLL 180

Query: 181 IGLGNGYNMVCSDAMAMIRETSEYMEIHDKELVIVKKDSVEVQDYDGNVIERGSYTAELD 240
 IGLG GYNMVCSDAMAMIRETSE+MEIHDKELVI+ KD V V DYDG + R SYTAELD
 Sbjct: 181 IGLGEGYNMVCSDAMAMIRETSEFMEIHDKELVILTKDKVTVDYDGKELIRDSYTAELD 240

Query: 241 LSDIGKGTYPFYMLKEIDEQPTVMRKLITYANESGDMNVDSDIKSVQEADRLYILAAG 300
 LSDIGKGTYPFYMLKEIDEQPTVMR+LISTYA+E+G++ VD II S+QEADRLYILAAG
 Sbjct: 241 LSDIGKGTYPFYMLKEIDEQPTVMRQLITYADETGNVQVDPALITSIQEADRLYILAAG 300

Query: 301 TSYHAGFAAKTMIEKLTDTTPVELGVSSEWGYNMPLLSKKPMFILLSQSGETADSRQVLVK 360
 TSYHAGFA K M+E+LTDTPVELGV+SEWGY+MPLLSSKKPMFILLSQSGETADSRQVLVK
 Sbjct: 301 TSYHAGFATKNMLEQLTDTTPVELGVASEWGYHMPLLSKKPMFILLSQSGETADSRQVLVK 360

Query: 361 ANEMGIPSLTITNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQVATLAFLAKAVGEA 420
 AN MGIPSLT+TNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQ+A LAFLAKAVGEA
 Sbjct: 361 ANAMGIPSLTITNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQIAALAFKAVGEA 420

Query: 421 NGKAEAKDFDLVHELSSIVAQSIEATLSEKDVISEKVEQLLISTRNAFYIGRGNDYYVTME 480
 NGK EA DF+LVHELSS+VAQSIEATLSEK+++EKV+ LL +TRNAFYIGRGNDYYV ME
 Sbjct: 421 NGKQEAALDFNLVHELSSIVAQSIEATLSEKDLVAEKVQALLATRNAFYIGRGNDYYVAME 480

Query: 481 AALKLKEISYIQTEGFAAGELKHGTISLIEDNTPVIALISADSTIAAHTRGNIQEVVSRG 540
 AALKLKEISYIQ EGFAAGELKHGTISLIE++TPVIALIS+ +A+HTRGNIQEV +RG
 Sbjct: 481 AALKLKEISYIQCEGFAAGELKHGTISLIEEDTPVIALISSQLVASHTRGNIQEVAAARG 540

Query: 541 ANALIIVEEGLEREGDDIIVNKVHPFLSAISMVPTQLIAYYASLQRLDVDKPRNLAKA 600
 A+ L +VEEGL+REGDDIIVNKVHPFL+ I+MVPTQLIAYYASLQRLDVDKPRNLAKA
 Sbjct: 541 AHVLTVEEGLDREGDDIIVNKVHPFLAPIAMVPTQLIAYYASLQRLDVDKPRNLAKA 600

Query: 601 VTVE 604
 VTVE
 Sbjct: 601 VTVE 604

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1331

A DNA sequence (GBSx1415) was identified in *S.agalactiae* <SEQ ID 4081> which encodes the amino acid sequence <SEQ ID 4082>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have a cleavable N-term signal seq.

-1467-

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9797> which encodes amino acid sequence <SEQ ID 9798> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44435 GB:U65000 type-I signal peptidase SpsB [Staphylococcus aureus]
 Identities = 62/185 (33%), Positives = 97/185 (51%), Gaps = 12/185 (6%)

Query: 10 VKRDFIRNIILALIAVLILILLRYFVFATFKVHKDATNSYFSNGDVVVVN---RNRTPK 65
 +K++ + II +A +IL ++ F+ + + ++ + +G+ V VN + +
 Sbjct: 1 MKKELLEWIISI AVAFVILFIVGKFIVTPYTIKGESMDPTLKDGERVAVNIIGYKTGGL 60

Query: 66 YKDFIVYKVGKIF-YISRVIGEPNQKVRVMDLILYLNDFVKDEPYIEKMKNAYSEKKDGQ 124
 + +V+ K Y+ RVIG P KV +D LY+N +DEPY+ N + K G
 Sbjct: 61 KGNVVVFHANKNDYVVKRVIGVPGDKVEYKNDTLYVNGKKQDEPYL----NYNLKHKQGD 116

Query: 125 MPFTSDFSVELT--TRNKESRVPKGSYLVLNDNRQNKNSRKFGLIKEKDIRGVITFKVY 182
 T F V+ L K + +PKG YLVL DNR+ DSR FGLI E I G ++F+ +
 Sbjct: 117 Y-ITGTFQVKDLPLANPKSNVPIKGYLVLDNREVS KDSRAFGGLIDEDQIVGKVSFRFW 175

Query: 183 PLSEF 187
 P SEF
 Sbjct: 176 PFSEF 180

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4083> which encodes the amino acid sequence <SEQ ID 4084>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -14.22 Transmembrane 10 - 26 (4 - 34)

----- Final Results -----

bacterial membrane --- Certainty=0.6689(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 99/185 (53%), Positives = 130/185 (69%)

Query: 9 MVKRDIFIRNIILALIAVLILILLRYFVFATFKVHKDATNSYFSNGDVVVVNRNRTPKYKD 68
 MVKRDIFIRNI+L LI ++ ILLR FVF+TFKV + N+Y +GD+V + +N PKYKD
 Sbjct: 1 MVKRDIFIRNILLLLIVIGAILLRIFVFTFKVSPETANTYTLKSGDLVTIKKNIQPKYKD 60

Query: 69 FIVYKVGKIFYISRVIGEPNQKVRVMDLILYLNDFVKDEPYIEKMKNAYSEKKDQMPFT 128
 F+VY+VGK Y+SRVI V MDDI YLN++ + + Y+EKM K Y +T
 Sbjct: 61 FVVYRVGKKDYVSRVIAVEGDSVTYMDDI FYLNNMVESQAYLEKMKAHYLNHAPFGTLYT 120

Query: 129 SDFSVELTTRNKESRVPKGSYLVLNDNRQNKNSRKFGLIKEKDIRGVITFKVYPLSEFG 188
 DF+V T+T +K +VPKG YL+LNDNR+N NDSR+FGLI I+G++TF+V PLS+FG
 Sbjct: 121 DDFTVATTITADKYQKVPKGYLLLLNDNRKNTNDSRRFGLINASQIKGLVTFRVLPLSDFG 180

Query: 189 FTASE 193
 F E
 Sbjct: 181 FVEVE 185

A related GBS gene <SEQ ID 8789> and protein <SEQ ID 8790> were also identified. Analysis of this protein sequence reveals the following:

-1468-

Lipop: Possible site: -1 Crend: 10

McG: Discrim Score: 10.13

GvH: Signal Score (-7.5): 0.45

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 0 value: 3.82 threshold: 0.0

PERIPHERAL Likelihood = 3.82 69

modified ALOM score: -1.26

*** Reasoning Step: 3

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

36.0/59.9% over 165aa

Bacillus caldolyticus

EGAD|24914| signal peptidase i Insert characterized

ORF00169(364 - 867 of 1179)

EGAD|24914|25718(15 - 180 of 182) signal peptidase i {Bacillus caldolyticus}

%Match = 11.9

%Identity = 35.9 %Similarity = 59.9

Matches = 60 Mismatches = 61 Conservative Sub.s = 40

```

312      342      372      402      432      462      483      510
L*KHDIMEKRLGVVMVKRDFIRNIILALIAVLILILLRYFVFATFKVHKDATNSYFSNGDVVVNR---NRTPKYK-DFI
      |  ::      ::|| :: || ||| : | : : :|::||: : | : | |
      VTKQKEKRGRWPWFVAVCVVATLRLFLVFSNYVVEGKSMPTLESGNLLIVNKLSDIGPIRRFDII
              10      20      30      40      50      60

537      567      597      627      657      687      717      747
VYKVGKIF-YISRVIGEPNQKVRVMDLILYLVNDFKDEPYIEKMKNAYSEKKDQMPFTSDFSVETLTRNKESRVPKGSY
| : | | : |||| | : : ||||: | |||: | : ||: | ||: | : |||| |
VFHANKKEDYVKRVIGLPGDRIAYKNDILYVNGKKVDPEYLRPYKQ---KLLDGRL--TGDFTLLEVT--GKTRVPPGCI
      80      90      100      110      120      130      140

777      807      837      867      897      927      957      987
LVLNDNRQNKNDNRKFLIKEKDIRGVITFKVYPLSEFGFTASE**KNGII*YHSFYVIKWLNRNIFF*DR*NF**RXN*
: || ||| : ||| ||: | | : | : : | |
FVLGDNRLSSWDSRHFGFVKINQIVGKVDFRYWPFKQFAFQF
      150      160      170      180

```

SEQ ID 8790 (GBS7) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 4; MW 46kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 4; MW 21kDa). The GBS7-His fusion product was purified (Figure 189, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 262), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1332

A DNA sequence (GBSx1416) was identified in *S.agalactiae* <SEQ ID 4085> which encodes the amino acid sequence <SEQ ID 4086>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

-1469-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1099(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9795> which encodes amino acid sequence <SEQ ID 9796> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF25804 GB:AF172173 pyruvate kinase [Streptococcus thermophilus]
 Identities = 413/500 (82%), Positives = 451/500 (89%)

Query: 1 MNKRVKIVATLGPVAVEFRGGKKFGESGYWGESLDVEASAEKIAQLIKEGANVFRFNFSHG 60
 MNKRVKIVATLGPVAVE RGGKKFGE GYW E LD +ASA+ IAQLI+EGANVFRFNFSHG
 Sbjct: 1 MNKRVKIVATLGPVAVEIRGGKKFGEDGYWSEKLDPDASAKNIAQLIEGANVFRFNFSHG 60

Query: 61 DHAEQGMATVRKAEEIAGQKVGFLLDTKGPEIRTELFEDGADFHSTTTGTLRVATKQ 120
 +HAEQG RM VR AE IAGQKVGFLLDTKGPEIRTELF E A ++Y TG ++R+ATKQ
 Sbjct: 61 NHAEQGERMDVVRMAESIAGQKVGFLLDTKGPEIRTELFEGDAKEYAYKTGEQIRIATKQ 120

Query: 121 GIKSTPEVIALNVAGGLDIFDDVEVGKQILVDDGKLGTLVFAKDKDTREFEVVENDGLI 180
 G+KST +VIALNVAG LDIFDDVEVGKQ+LVDDGKLG L V KD + REF V VENDG+I
 Sbjct: 121 GLKSTRDVIALNVAGALDIFDDVEVGKQVLVDDGKLGRLVVDKDAEKREFIVEVENDGII 180

Query: 181 GKQKGVNIPYTKIPFPALAERDNADIRFGLEQGLNFIAISFVRTAKDVNEVRAICEETGN 240
 KQKGVNIPYTKIPFPALAERDNADIRFGLEQG+NFAISFVRTAKDV EVRAICEETGN
 Sbjct: 181 AKQKGVNIPYTKIPFPALAERDNADIRFGLEQGINFIAISFVRTAKDVQEVRAICEETGN 240

Query: 241 GHVKLFAKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPFEMVPVYQKMIITKVNAAGK 300
 GHVKL AKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPFEMVPVYQKMIITKVNAAGK
 Sbjct: 241 GHVKLLAKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPFEMVPVYQKMIITKVNAAGK 300

Query: 301 AVITATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360
 V+TATNMLETMT+KPRATRSEVSDVFNAVIDGTDATMLSGESANG YPVESVRTMATI
 Sbjct: 301 IVVTATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGPYPVESVRTMATIH 360

Query: 361 KNAQTLLNEYGRILDSSAFFRNKTDVIASAVKDATHSMDIKLVVTTITETGNTARAISKFR 420
 KNAQTLL EYGRIL+SS F R++ T+V+ASAVKDAT+SM I+L+V +TE+GNTA I +R
 Sbjct: 361 KNAQTLLKEYGRILNSSTFDRSSNTEVVASAVKDATNSMHIQLIVALTESGNTASLIDTYR 420

Query: 421 PDADILAVTFDEKVRSLMINWGVIPVLADKPASTDDMFVAERVALEAGFVESGDNIVI 480
 P+ADI A+TFDE Q+SLM+NWGVIPV+ + P+STDDMFVAERVALE+G VESGDNIVI
 Sbjct: 421 PEADIWAITFDELTQKSLMLNWGVIPVVTETPSSTDDMFVAERVALESGLVESGDNIVI 480

Query: 481 VAGVPVGTGGTNTMRVRTVK 500
 VAGVPVG+G TNTMR+RTVK
 Sbjct: 481 VAGVPVGSNTNTMRIRTVK 500

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4087> which encodes the amino acid sequence <SEQ ID 4088>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0915(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 272-274

The protein has homology with the following sequences in the databases:

>GP:AAF25804 GB:AF172173 pyruvate kinase [Streptococcus thermophilus]
 Identities = 404/500 (80%), Positives = 457/500 (90%)

-1470-

Query: 1 MNKRVKIVATLGPAVEIRGGKKYGEDGYWAGQLDVEESAKKIAELIEAGANVFRFNFSHG 60
 MNKRVKIVATLGPAVEIRGGKK+GEDGYW+ +LD + SAK IA+LIE GANVFRFNFSHG
 Sbjct: 1 MNKRVKIVATLGPAVEIRGGKKFGEDGYWSEKLDPDASAKNIAQLIEAGANVFRFNFSHG 60

Query: 61 DHKEQGDRMATVRLAEETARQKVGFLDITKGPEMRTELFADDAKEFSYVTGEKIRVATTQ 120
 +H EQG+RM VR+AE IA QKVGFLDITKGPE+RTELF DAKE++Y TGE+IR+AT Q
 Sbjct: 61 NHAEQGERMDVVRMAESIAGQKVGFLDITKGPEIRTELFEGDAKEYAYKTGEQIRIATKQ 120

Query: 121 GIQSTRDVIALNVAGSLDIYDEVEVGHTILIDGKGLKVIDKDIATRQFIVEVENDGII 180
 G++STRDVIALNVAG+LDI+D+VEVG +L+DDGKGL+V+DKD R+FIVEVENDGII
 Sbjct: 121 GLKSTRDVIALNVAGALDI+D+VEVGKQVLVDGKGLRVDKDAEKREFIVEVENDGII 180

Query: 181 AKQKGVNIPNTKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDVEEVREICRETGN 240
 AKQKGVNIP TKIPFPALAEERDADIRFGLEQG+NFIAISFVRTAKDV+EVR IC ETGN
 Sbjct: 181 AKQKGVNIPYTKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDVQEVRAICEETGN 240

Query: 241 DHVQLFAKIENQQGIDNLDIEIEAADGIMIARGDMGIEVPFEMVPVQKMIITKVNAAGK 300
 HV+L AKIENQQGIDN+DEIEEAADGIMIARGDMGIEVPFEMVPV+QKMIITKVNAAGK
 Sbjct: 241 GHVKLLAKIENQQGIDNLDIEIEAADGIMIARGDMGIEVPFEMVPVQKMIITKVNAAGK 300

Query: 301 AVITATNMLETMTKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360
 V+TATNMLETMTKPRATRSEVSDVFNAVIDGTDATMLSGESANG YPVESVRTMATI
 Sbjct: 301 IVVTATNMLETMTKPRATRSEVSDVFNAVIDGTDATMLSGESANGPYPVESVRTMATIH 360

Query: 361 RNAQTLLNEYGRDSSAFPRNTKTDVIASAVKDATHSMDIKLVTTITETGNTARAIKFR 420
 +NAQTLL EYGR+SS F R++ T+V+ASAVKDAT+SM I+L+V +TE+GNTA I +R
 Sbjct: 361 KNAQTLLKEYGRDSSAFPRNTKTDVIASAVKDATNSMHIQLIVALTESGNTASLIDTYR 420

Query: 421 PDADILAVTFDEKQVQALMINWGVIPVLAEKPASTDDMFVEAERVAEAGLVQSGDNIVI 480
 P+ADI A+TFDE Q++LM+NWGVIPV+ E P+STDDMFVEAERVA+E+GLV+SGDNIVI
 Sbjct: 421 PEADIWATTFDELTKSLMLNWGVIPVVTETPSSTDDMFVEAERVAESGLVESGDNIVI 480

Query: 481 VAGVPVGTGGTNTMRVRTVK 500
 VAGVPVG+G TMTMR+RTVK
 Sbjct: 481 VAGVPVGSNTNTMRIRTVK 500

An alignment of the GAS and GBS proteins is shown below.

Identities = 440/500 (88%), Positives = 462/500 (92%)

Query: 1 MNKRVKIVATLGPAVEIRGGKKFGESGYWGESLDVEESAELIAQLIKEGANVFRFNFSHG 60
 MNKRVKIVATLGPAVE RGGKK+GE GYW LDVE SA+KIA+LI+ GANVFRFNFSHG
 Sbjct: 1 MNKRVKIVATLGPAVEIRGGKKYGEDGYWAGQLDVEESAKKIAELIEAGANVFRFNFSHG 60

Query: 61 DHAEQGMATVRKAEETAGQKVGFLDITKGPEIRTELFEDGADFHSTYTTGKLRVATKQ 120
 DH EQG RMATVR AEEIA QKVGFLDITKGPE+RTELF D A SY TG K+RVAT Q
 Sbjct: 61 DHKEQGDRMATVRLAEETARQKVGFLDITKGPEMRTELFADDAKEFSYVTGEKIRVATTQ 120

Query: 121 GIKSTPEVIALNVAGGLDIFDDVEVGKQILVDDGKGLTVFAKDKDTREFEVVENDGLI 180
 GI+ST +VIALNVAG LDI+D+VEVG IL+DDGKGL V KD TR+F V VENDG+I
 Sbjct: 121 GIQSTRDVIALNVAGSLDIYDEVEVGHTILIDGKGLKVIDKDIATRQFIVEVENDGII 180

Query: 181 GKQKGVNIPYTKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDVNEVRAICEETGN 240
 KQKGVNIP TKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDV EVR IC ETGN
 Sbjct: 181 AKQKGVNIPNTKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDVEEVREICRETGN 240

Query: 241 GHVKLFAKIENQQGIDNLDIEIEAADGIMIARGDMGIEVPFEMVPVQKMIITKVNAAGK 300
 HV+LFAKIENQQGIDN+DEIEEAADGIMIARGDMGIEVPFEMVPV+QKMIITKVNAAGK
 Sbjct: 241 DHVQLFAKIENQQGIDNLDIEIEAADGIMIARGDMGIEVPFEMVPVQKMIITKVNAAGK 300

Query: 301 AVITATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360
 AVITATNMLETMT+KPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID
 Sbjct: 301 AVITATNMLETMTKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360

Query: 361 KNAQTLLNEYGRDSSAFPRNTKTDVIASAVKDATHSMDIKLVTTITETGNTARAIKFR 420
 +NAQTLLNEYGRDSSAFPR NKT+VIASAVKDATHSMDIKLVTTITETGNTARAIKFR
 Sbjct: 361 RNAQTLLNEYGRDSSAFPRNTKTDVIASAVKDATHSMDIKLVTTITETGNTARAIKFR 420

-1471-

Query: 421 PDADILAVTFDEKVVQSLMINWGVIPVLADKPASTDDMFEVAERVALEAGFVESGDNIVI 480
 PDADILAVTFDEKVVQSLMINWGVIPVLA+KPASTDDMFEVAERVA+EAG V+SGDNIVI
 Sbjct: 421 PDADILAVTFDEKVVQSLMINWGVIPVLA+KPASTDDMFEVAERVALEAGLVQSGDNIVI 480

Query: 481 VAGVPVGTGGTNTMRVRTVK 500
 VAGVPVGTGGTNTMRVRTVK
 Sbjct: 481 VAGVPVGTGGTNTMRVRTVK 500

10 A related GBS gene <SEQ ID 8791> and protein <SEQ ID 8792> were also identified. Analysis of this protein sequence reveals the following:

Belongs to Glycolysis/gluconeogenesis pathway. Proteins belonging to this metabolic pathway have been experimentally detected on the surface of *Streptococcus*.

The protein has homology with the following sequences in the databases:

15 >GP|6708108|gb|AAF25804.1|AF172173_2|AF172173 pyruvate kinase
 {*Streptococcus thermophilus*}

Score = 821 bits (2098), Expect = 0.0
 Identities = 412/500 (82%), Positives = 450/500 (89%)

20 Query: 1 MNKRKIVATLGPVAVEFRGGKKFGESGYWGESLDVEASAQIAQLIKEGANVFRFNFSHG 60
 MNKRKIVATLGPVAVE RGGKKFGE GYW E LD +ASA+ IAQLI+EGANVFRFNFSHG
 Sbjct: 1 MNKRKIVATLGPVAVEIRGGKKFGEDGYWSEKLDPDASAKNIAQLIEGANVFRFNFSHG 60

25 Query: 61 DHAEQGMATVRKAEEIAGQKVGFLDLTKGPEIRTELFEDGADFHSTYTGTKLRVATKQ 120
 +HAEQG RM VR AE IAGQKVGFLDLTKGPEIRTELF E A ++Y TG ++R+ATKQ
 Sbjct: 61 NHAEQGERMDVVRMAESIAGQKVGFLDLTKGPEIRTELFEGDAKEYAYKTGEQIRIATKQ 120

30 Query: 121 GIKSTPEVIALNVAGGLDIFDDVEVGKQILVDDGKLGTLVFAKDKDTREFEVVENDGLI 180
 G+KST +VIALNVAG LDIFDDVEVGKQ+LVDDGKLG L V KD + REF V VENDG+I
 Sbjct: 121 GLKSTRDVIALNVAGALDIFDDVEVGKQVLVDDGKGLRVVDKDAEKREFIVEVENDGII 180

Query: 181 GKQKGVNIPYTKIPFPALAEARNADIRFGLEQGLNFIAISFVRTAKDVNEVRAICEETGX 240
 KQKGVNIPYTKIPFPALAEARNADIRFGLEQG+NFIAISFVRTAKDV EVRAICEETG
 35 Sbjct: 181 AKQKGVNIPYTKIPFPALAEARNADIRFGLEQGINFIAISFVRTAKDVQEVRAICEETGN 240

Query: 241 GHVKLFAKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPFEMVPVYQMIITKVNAAGK 300
 GHVKL AKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPFEMVPVYQMIITKVNAAGK
 Sbjct: 241 GHVKLLAKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPFEMVPVYQMIITKVNAAGK 300

40 Query: 301 AVITATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360
 V+TATNMLETMT+KPRATRSEVSDVFNAVIDGTDATMLSGESANG YPVESVRTMATI
 Sbjct: 301 IVVTATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGPYPVESVRTMATIH 360

45 Query: 361 KNAQTLLNEYGRLDSSAFPRNKTVDIASAVKQATHSMDIKLVVTTITETGNTARAISKFR 420
 KNAQTLL EYGRLLSS F R++ T+V+ASAVKDAT+SM I+L+V +TE+GNTA I +R
 Sbjct: 361 KNAQTLLKEYGRLLNSSTFDRSSNTEVVASAVKDATNSMHIQLIVALTESGNTASLIDTYR 420

50 Query: 421 PDADILAVTFDEKVVQSLMINWGVIPVLADKPASTDDMFEVAERVALEAGFVESGDNIVI 480
 P+ADI A+TFDE Q+SLM+NWGVIPV+ + P+STDDMFEVAERVALE+G VESGDNIVI
 Sbjct: 421 PEADIWATTFDELTQKSLMLNWGVIPVVTETPSSTDDMFEVAERVALESGLVESGDNIVI 480

Query: 481 VAGVPVGTGGTNTMRVRTVK 500
 VAGVPVG+G TMTMR+RTVK
 55 Sbjct: 481 VAGVPVGSNTNTMRIRTVK 500

SEQ ID 8792 (GBS330) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 5; MW 59kDa).

GBS330-His was purified as shown in Figure 213, lane 6.

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1472-

Example 1333

A DNA sequence (GBSx1417) was identified in *S.agalactiae* <SEQ ID 4089> which encodes the amino acid sequence <SEQ ID 4090>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0632(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF25803 GB:AF172173 phosphofructokinase [Streptococcus thermophilus]
Identities = 270/337 (80%), Positives = 302/337 (89%), Gaps = 1/337 (0%)

Query: 1 MKRIAVLTSGGDAPGMNAAIRAVVRKAISEGMEVYGINQGYGMVTGDIFPLDANSVGDT 60

MKRIAVLTSGGDAPGMNAA+RAVV KAISEG+EV+GIN+GY GMV GDIF LDA V +

Sbjct: 1 MKRIAVLTSGGDAPGMNAAVRAVLKAISEGIEVFGINRGYAGMVEGDIFKLDKRVENI 60

Query: 61 INRGGTFLRSARYPEFAELEGQLKGIEQLKKHGIEGVVIGDGSYHGAMRLTEHGFFAV 120

++RGGTFL+SARYPEFA+LEGQLKGIEQLKK+GIEGVVIGDGSYHGAMRLTEHGFFAV

Sbjct: 61 LSRGGTFLQSARYPEFAQLEGQLKGIEQLKKYIEGVVIGDGSYHGAMRLTEHGFFAV 120

Query: 121 GLPGTIDNDIVGTDYTGIFDTAVATAVENLDRLRDTSSASHNRTFVVEVMGRNAGDIALWS 180

GLPGTIDNDIVGTDYTGIFDTAVATA E LD+++DT+ SH RTFVVEVMGRNAGDIALW+

Sbjct: 121 GLPGTIDNDIVGTDYTGIFDTAVATATEALDKIQDTAFSHGRFTFVVEVMGRNAGDIALWA 180

Query: 181 GIAAGADQIIVPEEEFNIDEVSVNRAGYAAG-KHHQIIVLAEGVMGDEFKTMKAAGD 239

GIA+GADQIIVPEEE++I+EVV V+ GY +G K H IIVLAEGVM +EFA MK AGD

Sbjct: 181 GIASGADQIIVPEEYDINEVVRKVEGYESGEKSHHIIIVLAEGVMGAEEFAAKMKEAGD 240

Query: 240 DSDLRVNTNLGHLLRGGSPRTARDRVLASRMGAYAVQLLKEGRGGLAVGVHNEEMVESPILG 299

SDLR TNLGH++RGGSPRTARDRVLAS MGA+AV LLKEG GG+AVG+HNE++VESPILG

Sbjct: 241 TSDLRATNLGHVIRGGSPRTARDRVLASWMAHAVDLLKEGIGGVAVGIHNEQLVESPILG 300

Query: 300 LAEEGALFSLTDEGKIIVNNPHKADLRRLAALNRDLAN 336

AEEGALFSLT++GKI+VNNPHKA L A LNR LAN

Sbjct: 301 TAEAGALFSLTDEGKIIVNNPHKARLDFAEILNRSLAN 337

Proteins in the glycolysis/gluconeogenesis pathway have been experimentally detected on the surface of Streptococci.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4091> which encodes the amino acid sequence <SEQ ID 4092>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0632(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 274/336 (81%), Positives = 306/336 (90%), Gaps = 1/336 (0%)

Query: 1 MKRIAVLTSGGDAPGMNAAIRAVVRKAISEGMEVYGINQGYGMVTGDIFPLDANSVGDT 60

MKRIAVLTSGGDAPGMNAAIRAVVRKAISEGMEVYGIN+GY GMV GDIFPL + VGD

Sbjct: 1 MKRIAVLTSGGDAPGMNAAIRAVVRKAISEGMEVYGINRGYAGMVDGDTFPLGSKEVGDK 60

Query: 61 INRGGTFLRSARYPEFAELEGQLKGIEQLKKHGIEGVVIGDGSYHGAMRLTEHGFFAV 120

-1473-

```

I+RGGTFL SARYPEFA+LEGQL GIEQLKKGIEGVVIGDGSYHGAMRLTEHGFFAV
Sbjct: 61 ISRGGTFLYSARYPEFAQLEGQLAGIEQLKKGIEGVVIGDGSYHGAMRLTEHGFFAV 120

Query: 121 GLPGTIDNDIVGTDYTGIFDITAVATAVENLDRDRDTSASHNRTFVVEVMGRNAGDIALWS 180
G+PGTIDNDI GTDYTGIFDITAV TAVE +D+LRDTS+SH RTFVVEVMGRNAGDIALW+
Sbjct: 121 GIPGTIDNDIAGTDYTGIFDITAVTAVEAIDKLRDTSSSHGRTFVVEVMGRNAGDIALWA 180

Query: 181 GIAAGADQIIVPEEEFNIDEVVSNNRAGYA-AGKHHQIIVLAEGVMSGDEFKTMKAAGD 239
GIA+GADQIIVPEEEF+I++V S ++ + GK+H IIVLAEGVMSG+ FA+ +K AGD
Sbjct: 181 GIASGADQIIVPEEEFDIEKVASTIQYDFEHKGKNNHIIIVLAEGVMSGFAFAQKLKEAGD 240

Query: 240 DSDLRVTNLGHLLRGGSPTARDRVLASRMGAYAVQLLKEGRGGLAVGVHNEEMVESPILG 299
SDLRVTNLGH+LRGGSPTARDRV+AS MG++AV+LLK+G+GGLAVG+HNEE+VESPILG
Sbjct: 241 KSDLRVTNLGHILRGGSPTARDRVIASWMGSHAVELLKDGKGLAVGIHNEELVESPILG 300

Query: 300 LAEEGALFSLTDEGKIVNNPHKADLRILAALNRDLA 335
AEEGALFSLT+EGKI+VNNPHKA L AALNR L+
Sbjct: 301 TAEAGALFSLTEGKIIVNNPHKARLDFEALNRSLs 336

```

20 SEQ ID 4090 (GBS313) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 5; MW 41kDa).

GBS313-His was purified as shown in Figure 204, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1334

A DNA sequence (GBSx1418) was identified in *S.agalactiae* <SEQ ID 4093> which encodes the amino acid sequence <SEQ ID 4094>. This protein is predicted to be DNA polymerase III alpha subunit (dnaE). Analysis of this protein sequence reveals the following:

```

Possible site: 55
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1446(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

There is also homology to SEQ ID 4096.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1335

A DNA sequence (GBSx1419) was identified in *S.agalactiae* <SEQ ID 4097> which encodes the amino acid sequence <SEQ ID 4098>. This protein is predicted to be YHCF (farR). Analysis of this protein sequence reveals the following:

```

Possible site: 52
45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-1474-

>GP:BA04102 GB:AP001508 transcriptional regulator (GntR family)
[Bacillus halodurans]
Identities = 51/116 (43%), Positives = 79/116 (67%)

5 Query: 5 FNEKSPIYSQIAEHIKMQIVSQEIKSGDQLPTVRELAQEAGVNPNTMQRAFTELEREGMV 64
F+ PIY Q+AE +K QIV E++ G++LP+VR++ EA VNPNT+QR + ELE +V
Sbjct: 5 FHSSEPIYLQLAERVKRQIVRGELRLGEKLPVSRDMGIEANVNPNTVQRTYRELEGLKIV 64

10 Query: 65 FSQRTSGRFVTTEDNLLIGKIRQQVAKAELATFVNNMKKIGYKLDEITVALDHFIKE 120
S+R G FVTED ++ IR+Q+ + E++ FV M+++GY +EI L+ ++ E
Sbjct: 65 ESKRGQGTFTVEDEQVLQAIREQMKETEISHFVQGMREMGYSNENIQAGLESYLTE 120

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4099> which encodes the amino acid sequence <SEQ ID 4100>. Analysis of this protein sequence reveals the following:

15 Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2075(Affirmative) < succ>
20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 80/120 (66%), Positives = 100/120 (82%)

25 Query: 1 MAWEFNEKSPIYSQIAEHIKMQIVSQEIKSGDQLPTVRELAQEAGVNPNTMQRAFTELER 60
M+W+F EKSPIY+QIA+H+ MQI+SQEIKSGDQLPTVRE A+ AGVNPNTMQRAFTELER
Sbjct: 1 MSWKFEKSPIYQIAQHVMQIISQEIKSGDQLPTVREYAEIAGVNPNTMQRAFTELER 60

30 Query: 61 EGMVFSQRTSGRFVTTEDNLLIGKIRQQVAKAELATFVNNMKKIGYKLDEITVALDHFIKE 120
EGMV+SQRT+GRFVT+D LI + R++++A +EL +F+ NM K+G+ EI L F+KE
Sbjct: 61 EGMVYSQRTAGRFVTTDDQKLIARKRRELAISELESFITNMTKMGFSHTTEIIPVLTSFLKE 120

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
35 vaccines or diagnostics.

Example 1336

A DNA sequence (GBSx1420) was identified in *S.agalactiae* <SEQ ID 4101> which encodes the amino acid sequence <SEQ ID 4102>. This protein is predicted to be ABC transporter, ATP-binding protein (yhG). Analysis of this protein sequence reveals the following:

40 Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2757(Affirmative) < succ>
45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12735 GB:Z99108 similar to glycine betaine/L-proline
50 transport [Bacillus subtilis]
Identities = 87/228 (38%), Positives = 150/228 (65%), Gaps = 1/228 (0%)

Query: 5 LQLHHVTKKYHKHTAVNDVTVSIPTGKIIGLLGPNNGSGKTTIIMINGLLQPDKGDIVID 64
++L HV+KKY +HTAVNDV++++ +G+I GL+GPNNGSGK+T +KM+ GLL P G + +D
55 Sbjct: 3 IKLEHVSKKYGRHTAVNDVSITLSSGRIYGLIGPNNGSGKSTTLKMMAGLLFPTSGFVKVD 62

Query: 65 GYRPSVETKKIISYLPDTSYQENMKIKDVVTLFEDFYNDFDSKVAYQLFEDLNLNPRER 124
+ + E + +YL + + +KD+V ++ + DF ++ Y+L ++ LNP ++

-1475-

Sbjct: 63 EEQVTREMRVQTAYLTLEDMFYPHFTVKDMVNFYQSQFPDFHTEQVYKLLNEMQLNPEKK 122

Query: 125 LKNLSKGNKEKVQLILVMSRKARLYILDEPIGGVDPAARDYILKTIISNYSNDAS-VLIS 183
 +K LSKGN+ +++++L ++R+A + +LDEP G+DP RD I+ +++S + V+I+

5 Sbjct: 123 IKKLSKGNRGRLLKIVLALARRADVILLDEPFSGLDPMVRDSIVNSLVSYIDFEQQIVVIA 182

Query: 184 THLISDIEPILDEVIFLKEGEIDLQGNADDLREEHNCSIDALFRERFK 231
 TH I +IE +LDEVI L GE Q +D+RE+ S+ F+ + +

10 Sbjct: 183 THEIDEIETLLDEVIILANGEKVAQREVEDIREQEGMSVLQWFKSKME 230

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4103> which encodes the amino acid sequence <SEQ ID 4104>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1983(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

An alignment of the GAS and GBS proteins is shown below.

Identities = 171/231 (74%), Positives = 200/231 (86%)

25 Query: 1 MTQLQLLHHVTKKYHKHTAVNDVTVSIPGKIIGLLGPNSGKTTIIMINGLLQPDKGD 60
 M LLQLHHV+K Y + A++D+T++IP GKIIGLLGPNSGKTT+IK+INGLLQP+KG+

Sbjct: 1 MAHLLQLLHHVSKSYREKKAIDDLTITIPNGKIIGLLGPNSGKTTLIKLINGLLQPNKGE 60

Query: 61 IVIDGYRPSVETKKIISYLPDTSYLQENMKIKDVVTLFEDFYNDFDSKVAYQLFEDLNLN 120
 IVIDGYRP VETKKIISYLPDT+YL ENM+IKD++ F DFY+DFD A L DL L+

30 Sbjct: 61 IVIDGYRPCVETKKIISYLPDTTYLNENMRIKDMLEFFSDFYSDFDKSKATSLRLDLELD 120

Query: 121 PRRLKNLSKGNKEKVQLILVMSRKARLYILDEPIGGVDPAARDYILKTIISNYSNDASV 180
 P +R K LSKGNKEKVQLILVMSRKARLY+LDEPIGGVDPAARDYILKTII++Y +ASV

35 Sbjct: 121 PEDRFKTLNLSKGNKEKVQLILVMSRKARLYVLDEPIGGVDPAARDYILKTIINSYCENASV 180

Query: 181 LISTHLISDIEPILDEVIFLKEGEIDLQGNADDLREEHNCSIDALFRERFK 231
 +ISTHLISDIEPILDEVIFLK+G + L GNADDLR+E+ SID+LFRE +K

Sbjct: 181 IISTHLISDIEPILDEVIFLKQGRFLSLGNADDLRQEFYQQSIDSLEFRITYK 231

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1337

A DNA sequence (GBSx1421) was identified in *S.agalactiae* <SEQ ID 4105> which encodes the amino acid sequence <SEQ ID 4106>. Analysis of this protein sequence reveals the following:

45 Possible site: 48
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -15.39	Transmembrane	120 - 136 (103 - 146)
INTEGRAL	Likelihood = -9.98	Transmembrane	55 - 71 (47 - 79)
INTEGRAL	Likelihood = -9.45	Transmembrane	22 - 38 (15 - 43)
INTEGRAL	Likelihood = -6.05	Transmembrane	192 - 208 (187 - 218)
INTEGRAL	Likelihood = -4.94	Transmembrane	230 - 246 (228 - 253)
INTEGRAL	Likelihood = -4.78	Transmembrane	157 - 173 (155 - 175)
INTEGRAL	Likelihood = -1.44	Transmembrane	103 - 119 (103 - 119)

50 ----- Final Results -----

bacterial membrane --- Certainty=0.7156(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55

60 The protein has no significant homology with any sequences in the GENPEPT database.

-1476-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4107> which encodes the amino acid sequence <SEQ ID 4108>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

```

5      INTEGRAL    Likelihood = -11.52    Transmembrane  190 - 206 ( 187 - 215)
      INTEGRAL    Likelihood = -10.67    Transmembrane  121 - 137 ( 104 - 141)
      INTEGRAL    Likelihood = -5.73     Transmembrane   63 -  79 (  59 -  82)
      INTEGRAL    Likelihood = -4.83     Transmembrane  158 - 174 ( 156 - 181)
10     INTEGRAL    Likelihood = -1.38     Transmembrane  232 - 248 ( 232 - 248)
      INTEGRAL    Likelihood = -0.85     Transmembrane  104 - 120 ( 104 - 120)

```

----- Final Results -----

```

      bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
15     bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

20     Identities = 116/267 (43%), Positives = 165/267 (61%), Gaps = 13/267 (4%)

Query: 1  MFGKLLKYELKSVGKWYLTFLNAAVLLVSIILGLVLKALG-----GNFSTDNTNSTSAQIFT 55
      MFGKLLKYE +S+GKWY LNA V+ ++ IL +K G F TN ++
Sbjct: 1  MFGKLLKYEFRSIGKWYFALNAFVIAIAAILSFTIKLFAQSNSDGLFGVLTN----KMLP 56

25     Query: 56  IILVLLLAMVISGSLSTLAIIRFYSNIFGRQGYLTTLTPVTTNQTICSKLLASLLWS 115
      + L L +I+GSLSTL IIRKF ++FG +GYLTTLTPV ++QII SKLLAS + S
Sbjct: 57  LTLGLTFGSLIAGSLLSTLLIIIRFSKSVFGWEGYLTTLTPVNSHQIILSKLLASFICS 116

30     Query: 116 IFNIFIVIIGIILVILPLVGIGQFVVAFFPEIYKIISSSNAPLFIAFYFLLSYVAGTLLIYL 175
      +FN I+ I +VI+P+ I + + F +K+ N +AY LS LLIYL
Sbjct: 117 VFNTIILAFIAIVIVPMFNINELLEGGFFNSFKMDYFINMLTVLAYVLLSTFTSILLIYL 176

35     Query: 176 SIAVGQLFTNKRVLMGIVSYFGISLLITFLTLIIDSIFHIDLFNSHANA-TFSQPVLly- 233
      SI++GQLF+N+R LM ++YF + +LI+ + S HI N+ A++ F++ +Y
Sbjct: 177 SISIGQLFSNRRGLMAFIAYFILVILISVAATYVHS--HIFNINTSADSFPFTEQKTIYL 234

40     Query: 234 NILVSIVEIAIFYMLTHSIIKYKLNQ 260
      IL +E+ +FY+ T+ IIK KLN+Q
Sbjct: 235 LILEQFIEMIMFYLATNFIKKNLNQ 261

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1338

45 A DNA sequence (GBSx1422) was identified in *S.agalactiae* <SEQ ID 4109> which encodes the amino acid sequence <SEQ ID 4110>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

50     bacterial cytoplasm --- Certainty=0.5890(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein is similar to ORF24 from *S.faecalis*.

55 No corresponding DNA sequence was identified in *S.pyogenes*.

-1477-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1339

A DNA sequence (GBSx1423) was identified in *S.agalactiae* <SEQ ID 4111> which encodes the amino acid sequence <SEQ ID 4112>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to ORF23 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1340

A DNA sequence (GBSx1424) was identified in *S.agalactiae* <SEQ ID 4113> which encodes the amino acid sequence <SEQ ID 4114>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.4256(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to ORF22 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1341

A DNA sequence (GBSx1425) was identified in *S.agalactiae* <SEQ ID 4115> which encodes the amino acid sequence <SEQ ID 4116>. Analysis of this protein sequence reveals the following:

Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood =-13.37 Transmembrane 62 - 78 (55 - 84)
INTEGRAL Likelihood = -8.44 Transmembrane 19 - 35 (14 - 41)

----- Final Results -----

40 bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to ORF21 from *S.faecalis*.

-1478-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4117> which encodes the amino acid sequence <SEQ ID 4118>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2444(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 54/236 (22%), Positives = 95/236 (39%), Gaps = 12/236 (5%)

Query: 204 KDGKLRMLKNVWWEYDKLPMLIAGGTGGGKTYFILTLIEALLHTDSKLYILDPKN---- 259

+ GK+ ++K+ DK H IAG +G GK Y LT ++L S L I+DPK

Sbjct: 14 QQKIPVIKHFELNLDKGSHWAIAGNSGSGKPY-ALTYFLSVLPKPSGLIIDPKFDTPS 72

Query: 260 --ADLADLGSVMANVYYRKEDLLSCIETFFYEEMMKRSEEMKQMKNYKTGKNYAYLGLPAH 317

A + + + K D +S + + ++ + + + +L +

Sbjct: 73 QWARENKIAVIHPVENHKSDFVSQVNEQLNQCATLIQKRQAILYDNPNHQFTHLTI--- 129

Query: 318 FLIFDEYVAFMEMLGTKENTAVMNKLKQIVMLGRQAGFFLILACQRPDAKYLGDGIRDQF 377

+ DE +A E + A + L QI +LG L L QR D + +R+Q

Sbjct: 130 --VIDEVLALSEGYNKNIKEAFFSLLSQIALLGHATKIHLFLGSQRFDHNTIPISVREQL 187

Query: 378 NFRVALGRMSEMGMFSGSDVQKDFFLKRIKGRGYVDVGTSEFYTPPLVPKGY 433

N + +G +++ +F + + G G + V + S PL+ Y

Sbjct: 188 NVLLQIGNINQKTTQFLFPDLDPGIVIPTGHGTGIIQVVDNEHSYQVLPLLCPY 243

SEQ ID 4116 (GBS109d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 8 & 9; MW 71kDa) and in Figure 184 (lane 2; MW 71kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 11; MW 46kDa), Figure 128 (lane 4; MW 46kDa) and Figure 179 (lane 7; MW 46kDa). GBS109d-His was purified as shown in Figure 232 (lanes 7 & 8). GBS109d-GST was purified as shown in Figure 236, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1342

A DNA sequence (GBSx1426) was identified in *S.agalactiae* <SEQ ID 4119> which encodes the amino acid sequence <SEQ ID 4120>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1479-

Example 1343

A DNA sequence (GBSx1427) was identified in *S.agalactiae* <SEQ ID 4121> which encodes the amino acid sequence <SEQ ID 4122>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4469(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9793> which encodes amino acid sequence <SEQ ID 9794> was also identified.

The protein is similar to ORF20 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1344

A DNA sequence (GBSx1428) was identified in *S.agalactiae* <SEQ ID 4123> which encodes the amino acid sequence <SEQ ID 4124>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1367(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1345

A DNA sequence (GBSx1429) was identified in *S.agalactiae* <SEQ ID 4125> which encodes the amino acid sequence <SEQ ID 4126>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-10.77 Transmembrane 39 - 55 (34 - 64)

INTEGRAL Likelihood = -6.32 Transmembrane 16 - 32 (10 - 35)

----- Final Results -----

bacterial membrane --- Certainty=0.5310(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to ORF19 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

-1480-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1346

A DNA sequence (GBSx1430) was identified in *S.agalactiae* <SEQ ID 4127> which encodes the amino acid sequence <SEQ ID 4128>. This protein is predicted to be antirestriction protein. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
```

```
    bacterial cytoplasm --- Certainty=0.2918(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein is similar to ORF18 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1347

A DNA sequence (GBSx1431) was identified in *S.agalactiae* <SEQ ID 4129> which encodes the amino acid sequence <SEQ ID 4130>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -3.61    Transmembrane    75 - 91 ( 72 - 94)
```

```
----- Final Results -----
```

```
    bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein is similar to ORF17 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8793> and protein <SEQ ID 8794> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1    Crend: 4
McG: Discrim Score:      -7.12
GvH: Signal Score (-7.5): -2.52
    Possible site: 43
>>> Seems to have no N-terminal signal sequence
ALOM program    count: 1 value: -3.61 threshold: 0.0
    INTEGRAL    Likelihood = -3.61    Transmembrane    37 - 53 ( 34 - 56)
    PERIPHERAL  Likelihood =  3.66      58
    modified ALOM score:  1.22
```

```
*** Reasoning Step: 3
```

```
----- Final Results -----
```

```
    bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

-1481-

100.0/100.0% over 167aa

Enterococcus faecalis

EGAD|14977| hypothetical protein Insert characterized
 GP|532550|gb|AAB60016.1||U09422 ORF17 Insert characterized

ORF00720(187 - 690 of 990)

EGAD|14977|15011(1 - 168 of 168) hypothetical protein {Enterococcus faecalis}

GP|532550|gb|AAB60016.1||U09422 ORF17 {Enterococcus faecalis}

%Match = 50.3

%Identity = 100.0 %Similarity = 100.0

Matches = 168 Mismatches = 0 Conservative Sub.s = 0

```

120      150      180      210      240      270      300      330
L*AKYQLVFKTILIIKPMVGI*TFQERLSQPIMGFLKSSIKSVGTLILLADFLFYGVAQSATPIFYERIDYMKKIRSYTSI
|||||
MGFLKSSIKSVGTLILLADFLFYGVAQSATPIFYERIDYMKKIRSYTSI
                        10      20      30      40

360      390      420      450      480      510      540      570
WSVEKVLVSINDFRLPPFITFTQMTWVFVSLFAVMILGNLPPLSMIEGAFLKYFGIPVAFTWFMSTKTFDGKKPYGFLKS
|||||
WSVEKVLVSINDFRLPPFITFTQMTWVFVSLFAVMILGNLPPLSMIEGAFLKYFGIPVAFTWFMSTKTFDGKKPYGFLKS
                        60      70      80      90      100      110      120

600      630      660      690      720      750      780      810
VIAYALRPKLTYAGKKVTLGRNQPEAITAVRSEFYGISN*IH*KQSRLE*RRGMLCLL*ACSLQLLISKSRTESENTSA*F
|||||
VIAYALRPKLTYAGKKVTLGRNQPEAITAVRSEFYGISN
                        140      150      160

```

SEQ ID 8794 (GBS223) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 7; MW 18kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1348

A DNA sequence (GBSx1432) was identified in *S.agalactiae* <SEQ ID 4131> which encodes the amino acid sequence <SEQ ID 4132>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4292(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9791> which encodes amino acid sequence <SEQ ID 9792> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1349

A DNA sequence (GBSx1433) was identified in *S.agalactiae* <SEQ ID 4133> which encodes the amino acid sequence <SEQ ID 4134>. Analysis of this protein sequence reveals the following:

-1482-

Possible site: 16

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -6.21 Transmembrane 350 - 366 (345 - 368)

INTEGRAL Likelihood = -0.32 Transmembrane 171 - 187 (171 - 188)

----- Final Results -----

bacterial membrane --- Certainty=0.3484(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1350

A DNA sequence (GBSx1434) was identified in *S.agalactiae* <SEQ ID 4135> which encodes the amino acid sequence <SEQ ID 4136>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -10.30 Transmembrane 154 - 170 (148 - 177)

INTEGRAL Likelihood = -10.30 Transmembrane 21 - 37 (17 - 50)

INTEGRAL Likelihood = -10.03 Transmembrane 320 - 336 (316 - 367)

INTEGRAL Likelihood = -7.43 Transmembrane 346 - 362 (337 - 367)

INTEGRAL Likelihood = -7.01 Transmembrane 186 - 202 (180 - 206)

INTEGRAL Likelihood = -5.36 Transmembrane 411 - 427 (404 - 430)

INTEGRAL Likelihood = -1.17 Transmembrane 386 - 402 (386 - 402)

----- Final Results -----

bacterial membrane --- Certainty=0.5118(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1351

A DNA sequence (GBSx1436) was identified in *S.agalactiae* <SEQ ID 4137> which encodes the amino acid sequence <SEQ ID 4138>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.6306(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1483-

Example 1352

A DNA sequence (GBSx1437) was identified in *S.agalactiae* <SEQ ID 4139> which encodes the amino acid sequence <SEQ ID 4140>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 22
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2973(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1353

A DNA sequence (GBSx1438) was identified in *S.agalactiae* <SEQ ID 4141> which encodes the amino acid sequence <SEQ ID 4142>. Analysis of this protein sequence reveals the following:

```

20     Possible site: 42
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3382(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

There is also homology to SEQ ID 4144.

A related GBS gene <SEQ ID 8795> and protein <SEQ ID 8796> were also identified. Analysis of this protein sequence reveals the following:

```

30     Lipop: Possible site: -1   Crend: 3
      McG: Discrim Score:      11.12
      GvH: Signal Score (-7.5): 0.27
      Possible site: 24
      >>> Seems to have a cleavable N-term signal seq.
35     ALOM program count: 0 value: 4.19 threshold: 0.0
      PERIPHERAL Likelihood = 4.19      69
      modified ALOM score: -1.34

      *** Reasoning Step: 3

40     ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

100.0/100.0% over 332aa
                                     Enterococcus faecalis
      EGAD|36209| hypothetical protein Insert characterized
50     GP|532547|gb|AAB60019.1||U09422 ORF14 Insert characterized

      ORF00727(301 - 1299 of 1599)
      EGAD|36209|37602(1 - 333 of 333) hypothetical protein {Enterococcus
      faecalis}GP|532547|gb|AAB60019.1||U09422 ORF14 {Enterococcus faecalis}

```

-1484-

%Match = 61.7
 %Identity = 100.0 %Similarity = 100.0
 Matches = 333 Mismatches = 0 Conservative Sub.s = 0

```

5      249      279      309      339      369      399      429      459
      CSKSTTTKYK*TTNQRHH*ESR*ETMKLKTLLVIGGSGFLMVFSLLLFVAILFSDEQDSGISNIHYGGVNVSAEVLAH
      "
      MKLKTLLVIGGSGFLMVFSLLLFVAILFSDEQDSGISNIHYGGVNVSAEVLAH
      10      20      30      40      50

10     489      519      549      579      609      639      669      699
      KPMVEKYAKEYGVEEYVNILLAI IQVESGGTAEDVMQSSSLGLPPNSLSTEESIKQGVKYFSELLASSERLSVDLESVI
      |||
      KPMVEKYAKEYGVEEYVNILLAI IQVESGGTAEDVMQSSSLGLPPNSLSTEESIKQGVKYFSELLASSERLSVDLESVI

15     70      80      90      100      110      120      130

      729      759      789      819      849      879      909      939
      QSYNYGGGFLGYVANRGNKYTFELAQSFKEYSGGEKVSYPNPPIAIPINGGWRNYNGNMFYVQLVTQYLVTTFDDDTVQ
      |||
      QSYNYGGGFLGYVANRGNKYTFELAQSFKEYSGGEKVSYPNPPIAIPINGGWRNYNGNMFYVQLVTQYLVTTFDDDTVQ
      150      160      170      180      190      200      210

      969      999      1029      1059      1089      1119      1149      1179
      AIMDEALKYEGWRYVYGGASPTTSFDCSGLTQWTYKGAGINLPRTAQQQYDVTQHIPLSEAQAGDLVFFHSTYNAGSYIT
      |||
      AIMDEALKYEGWRYVYGGASPTTSFDCSGLTQWTYKGAGINLPRTAQQQYDVTQHIPLSEAQAGDLVFFHSTYNAGSYIT
      230      240      250      260      270      280      290

      1209      1239      1269      1299      1329      1359      1389      1419
      HVGIYLGNNRMFHAGDPIGYADLTSPYWQQHLVGAGRIKQ*ERKI***NLEKIRIKKNRYQRKNLVSIRSILIKRL*LP
      |||
      HVGIYLGNNRMFHAGDPIGYADLTSPYWQQHLVGAGRIKQ
      310      320      330
  
```

35 SEQ ID 8796 (GBS155) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 10; MW 38kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 7; MW 62kDa).

The GBS155-GST fusion product was purified (Figure 111; see also Figure 198, lane 74) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1354

45 A DNA sequence (GBSx1439) was identified in *S.agalactiae* <SEQ ID 4145> which encodes the amino acid sequence <SEQ ID 4146>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -8.60      Transmembrane      37 - 53 ( 35 - 55)

50     ----- Final Results -----
           bacterial membrane --- Certainty=0.4439(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

55

-1485-

A related GBS nucleic acid sequence <SEQ ID 9789> which encodes amino acid sequence <SEQ ID 9790> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1355

A DNA sequence (GBSx1440) was identified in *S.agalactiae* <SEQ ID 4147> which encodes the amino acid sequence <SEQ ID 4148>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -0.00 Transmembrane 391 - 407 (391 - 407)

----- Final Results -----

bacterial membrane --- Certainty=0.1001(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9787> which encodes amino acid sequence <SEQ ID 9788> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4149> which encodes the amino acid sequence <SEQ ID 4150>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2027(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 183/669 (27%), Positives = 305/669 (45%), Gaps = 63/669 (9%)

Query: 7 KIINIGVLAHVDAGKTTLTESLLYNSGAITELGSVDKGTTRTDNTLLERQRGITITQTGIT 66
K NIG++AHVDAGKTT TE +LY +G I ++G +G ++ D E++RGITI + T

Sbjct: 9 KTRNIGIMAHVDAGKTTTTERILYYTGKIHKIGETHEGASQMDWMBEQERGITITSAA 68

Query: 67 SFQWENTKVNIIIDTPGHMDFLAEVYRSLSVLDGAILLISAKDGVQAQTRILFHALRKMGI 126
+ QW+ +VNIIDTPGH+DF EV RSL VLDGA+ ++ ++ GV+ QT ++ + G+

Sbjct: 69 TAQWDGHRVNIIDTPGHVDFTIEVQSRSLRVLDGAVTVLDSQSGVEPQTETVWRQATEYGV 128

Query: 127 PTIFFINKIDQNGIDLSTVYQDIKEKLSAEI-----VIKQKVELYPN 168
P I F NK+D+ G D Q + ++L A +IK K E+Y N

Sbjct: 129 PRIVFANKMDKIGADFLYSVQTLHDRLQANAHPIQLPIGAEDDFRGIIDLIKMAEIYTN 188

Query: 169 MCVTNFTES---EQW-----DTVIEGNDLLLEKYSKGKSLEALELEQEEESIRF 213
T+ E E++ + V E ++DL+ KY+ G+ + EL

Sbjct: 189 DLGTDILEEDIPEEYLEQAQEYREKLI EAVAETDEDLMMKYLEGEEITNDELIAIGIRKAT 248

Query: 214 HNCSLFPVYHGSAKNNIGIDNLEIVI-----TNKFYSSTHRGPSE---L 254
N FPV GSA N G+ +++ + N + P+

Sbjct: 249 INVEFFPVLGSAFKNKGVLMLDAVIAYLPSPLDIPAIGKGVNPDDAEEERPASDEEPF 308

Query: 255 CGNVFKIEYTKKRQRLAYIRLYSGVLHLRDSVRVSEKEKI----KVTEMYTSINGELCKI 310
FKI RL + R+YSGVL+ V + K K ++ +M+ + E I

Sbjct: 309 AALAFKIMTDPFVGRLTFFRVYSGVLNNGSYVMNTSKGKRERIGRILQMHANSRQE---I 365

-1486-

5 Query: 311 DRAYSGEIVILQN-EFLKINSVLGDTKLLPQRKKIENPHPLQLTTVEPSKPEQREMLLDA 369
 + Y+G+I + L D K + IE P P++Q VEP ++ + A
 Sbjct: 366 ETVYAGDIAAAVGLKDTTGTGSLTDEKAKVILESIEVPEPVIQLMVEPKSKADQDKMGVA 425

10 Query: 370 LLEISDSDFLLRYYVDSTTHEIILSFLGKVQMEVISALLQEKYHVEIELKEPTVIYME-- 427
 L ++++ DP R + T E +++ +G++ ++V+ ++ ++ VE + P V Y E
 Sbjct: 426 LQKLAEEEDPTFRVETINVETGETVIAGMGELHLDVLVDRMKREFKVEANVGAPQVSYRETF 485

15 Query: 488 YGCEQG-LYGWNVTDCKICFKYGLYSPVSTPADFRMLAPIVLEQVLKKAGTELLEPYLS 546
 G L G+ + D K G Y+ S+ F++ A + L++ K A +LEP +
 Sbjct: 546 ESMANGVLAGYPMVDVKAKLYDGSYHDVDSSETAFKIAASLALKEAKSAQPAILEPMLI 605

20 Query: 547 FKIYAPQEYLSRAYNDAPKYCANIVDTQLKNNEVILSGEIPARCIQEYRSDLTFFTNGRS 606
 I AP++ L + + N I+ +P + Y + L T GR
 Sbjct: 606 VTITAPEDNLGDVMGHVTARRGRVDGMEAHGNSQIVRAYVPLAEMFGYATVLRSATQGRG 665

25 Query: 607 VCLTELKGY 615
 + Y
 Sbjct: 666 TFMMVFDHY 674

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1356

30 A DNA sequence (GBSx1441) was identified in *S.agalactiae* <SEQ ID 4151> which encodes the amino acid sequence <SEQ ID 4152>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2530(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1357

A DNA sequence (GBSx1442) was identified in *S.agalactiae* <SEQ ID 4153> which encodes the amino acid sequence <SEQ ID 4154>. Analysis of this protein sequence reveals the following:

45 Possible site: 18
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1487-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1358

A DNA sequence (GBSx1443) was identified in *S.agalactiae* <SEQ ID 4155> which encodes the amino acid sequence <SEQ ID 4156>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1630(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1359

A DNA sequence (GBSx1444) was identified in *S.agalactiae* <SEQ ID 4157> which encodes the amino acid sequence <SEQ ID 4158>. This protein is predicted to be excisionase-related protein. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4481(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to transposon Tn916 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1360

A DNA sequence (GBSx1445) was identified in *S.agalactiae* <SEQ ID 4159> which encodes the amino acid sequence <SEQ ID 4160>. This protein is predicted to be transposase. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4626(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar the Tn1545 integrase from *S.pneumoniae* and to SEQ ID 578.

-1488-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1361

A DNA sequence (GBSx1446) was identified in *S.agalactiae* <SEQ ID 4161> which encodes the amino acid sequence <SEQ ID 4162>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.72	Transmembrane	18 - 34 (13 - 41)
INTEGRAL	Likelihood = -6.10	Transmembrane	58 - 74 (55 - 79)
INTEGRAL	Likelihood = -5.04	Transmembrane	97 - 113 (90 - 116)
INTEGRAL	Likelihood = -1.81	Transmembrane	78 - 94 (78 - 94)
INTEGRAL	Likelihood = -0.85	Transmembrane	145 - 161 (145 - 161)

----- Final Results -----

bacterial membrane	---	Certainty=0.5288(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC74820 GB:AE000270 orf, hypothetical protein [Escherichia coli K12]
Identities = 43/174 (24%), Positives = 84/174 (47%), Gaps = 9/174 (5%)

Query: 24 LIATLVLVVYLYKL-----GILNDSNELKDLVHKYEFWGPIMIFIVAQIVQIVFPVIPGG 77
L A L+ + +Y + +L D L+ L+ + F+G ++I+ I+ + ++PG

Sbjct: 24 LFACLIFALVIYAIHAFGLFDLLTDLPHLQTLIRQSGFFGYSLYILLFIIATLL-LLPGS 82

Query: 78 VTTVAGFLIFGPTLGFYIYIGIIGSVILFVLVKFYGRKFVLLFM-DQKTFDKYESKLE 136
+ +AG ++FGP LG + + I + S F L ++ GR +L ++ TF E +

Sbjct: 83 ILVIAGGIVFGPLLGLTLLSLIAATLASSCSFILLARWLGRDLLKLVGHSTNTFQAIEKGIA 142

Query: 137 TSGYEKFFIFCMASPIPADIMVMITGLSNMSIKRFVTIIMITKPIISIIGYSYL 190

+G + F I P+ P +I GL+ ++ + I +T I+ Y+ +
Sbjct: 143 RNGID-FLILTRLIPLFPYNIQNYAYGLTTIAFWPYTLISALTLPGIVIIYTV 195

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4163> which encodes the amino acid sequence <SEQ ID 4164>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -4.30	Transmembrane	8 - 24 (6 - 29)
INTEGRAL	Likelihood = -0.80	Transmembrane	57 - 73 (57 - 73)
INTEGRAL	Likelihood = -0.00	Transmembrane	86 - 102 (86 - 102)

----- Final Results -----

bacterial membrane	---	Certainty=0.2720(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 85/114 (74%), Positives = 101/114 (88%)

Query: 89 PTLGFYIYNYIGIIGSVILFVLVKFYGRKFVLLFMDQKTFDKYESKLETSGYEKFFIFCM 148

P GFYIYNY+GIIIGS+ LF LVK YGRKF+LLF++ KTF KYE +LET GYEK FIFCM

Sbjct: 3 PVTGFYIYNYVGIIIGSIALFLLVKTGYGRKFILLFVNDKTFYKYERRLETPGYEKLIFIFCM 62

Query: 149 ASPISPADIMVMITGLSNMSIKRFVTIIMITKPIISIIGYSYLWIYGGDILKNFL 202

ASP+SPADIMVMITGL++MS+KRFVTI++ITKPIISIIGYSYL+I+G D++ FL

Sbjct: 63 ASPVSPADIMVMITGLTDMSLKRFVTIILLITKPIISIIGYSYLFIFGKDVISWFL 116

-1489-

There is also homology to SEQ ID 1728.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1362

A DNA sequence (GBSx1447) was identified in *S.agalactiae* <SEQ ID 4165> which encodes the amino acid sequence <SEQ ID 4166>. This protein is predicted to be chlorAMPhenicol acetyltransferase (cat). Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4725(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA86871 GB:U19459 VAT B [Staphylococcus aureus]
Identities = 57/130 (43%), Positives = 81/130 (61%), Gaps = 4/130 (3%)

20 Query: 57 IGAFCSIAQNVT--ITGLNHPTDHITTNPFYIYKSRGFINEADRADLIDEKNGKVIIGND 114
      IG FC+IA+ + + G NH + ITT PF G+ + L D G ++GND
      Sbjct: 65 IGKFCIAEGIEFIMNGANHRMNSITTYPF-NIMGNGW-EKATPSLEDLPFKGDTVVGND 122

25 Query: 115 VWIGTNVTILPSVTIGNGAIIAGSVITKDIPDYAVVAGTPAKIIKYRFSEEEITILNAS 174
      VWIG NVT++P + IG+GAI+ A SV+TKD+P Y ++ G P++IIK RF +E I L
      Sbjct: 123 VWIGQNVTVMPGIQIGDGAIVAANSVVTKDVPYRIIGGNPSRIIKRFEDELIDYLLQI 182

30 Query: 175 QWWNWSDEAI 184
      +WW+WS + I
      Sbjct: 183 KWWDSWAQKI 192

```

There is also homology to SEQ ID 1944.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1363

A DNA sequence (GBSx1448) was identified in *S.agalactiae* <SEQ ID 4167> which encodes the amino acid sequence <SEQ ID 4168>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2398(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1490-

Example 1364

A DNA sequence (GBSx1449) was identified in *S.agalactiae* <SEQ ID 4169> which encodes the amino acid sequence <SEQ ID 4170>. This protein is predicted to be cation-transporting P-ATPase PacL. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.18    Transmembrane  873 - 889 ( 866 - 894)
      INTEGRAL    Likelihood = -8.39    Transmembrane  257 - 273 ( 251 - 276)
      INTEGRAL    Likelihood = -5.95    Transmembrane   67 -  83 (  65 -  88)
10   INTEGRAL    Likelihood = -5.41    Transmembrane  282 - 298 ( 281 - 301)
      INTEGRAL    Likelihood = -1.65    Transmembrane   90 - 106 (  89 - 107)
      INTEGRAL    Likelihood = -0.48    Transmembrane  737 - 753 ( 736 - 753)
      INTEGRAL    Likelihood = -0.00    Transmembrane  898 - 914 ( 898 - 914)

15   ----- Final Results -----
      bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

20 A related GBS nucleic acid sequence <SEQ ID 10963> which encodes amino acid sequence <SEQ ID 10964> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAB85991 GB:AE000912 cation-transporting P-ATPase PacL
[Methanothermobacter thermoautotrophicus]
25   Identities = 409/922 (44%), Positives = 609/922 (65%), Gaps = 22/922 (2%)

Query: 10  TNTRFAKEELEEVFEELGTTQGGSLDEEVAVRQKKYGLNLLSEVKQESIIILLFLKNFTSL 69
      T T  + E+EEV + L T++ GL +E  R K +G N L EVK+ +ILLFL N ++
Sbjct: 4   TMTAIYELEVEEVLQRLETSESGLDPOEAEKRLKIHGPNKLEEVKRRPLILLFLSNLYNV 63

30   Query: 70  MAILLWVGGFVAIVSNSLELGLAIWMVNVINGIFSFIEYRASQATQALEKMLPSYSRVL 129
      +A+LLW+ ++ ++ + +L +AI MV +IN +FSF QEY A +A +AL+ +LP +V+
Sbjct: 64  LALLLWIAAILSFITGNYQLAVAIVMVIINALFSFWQEYEAKEAAEALKNILPVMVKVI 123

35   Query: 130 RKGSEEEKILSEQLVPGDIVLIEEGDRISADGRLIKTTDLQVNQSALTGESNPIYKDSNVE 189
      R  E  I + +V GDI+++EEGD + AD R++++ +L+V+ SALTGES P+ K S+
Sbjct: 124 RASKEVLIPAADVVHGDIIILEEGDTPADARILESHNLRVDASALTGESKPVKRVSHPV 183

40   Query: 190 NDQSKTLIECDNMVVFAGTTVSSGSATMVVTAIGMQTQFGQIADLTQGMKSEKSPLORELD 249
      + +  I+ +N++FAGT V+SG+  V A G T+F +IA LTQ ++ E SPLQR++
Sbjct: 184 RE-ADNYIDTENILFAGTQVTSGTGRAAVFATGRDTEFSRIATLTQEVREEPSPLQRQIS 242

Query: 250  RLTKQISIIISITVGIIFFLAATFFVKEPVSFSFIFALGMIVAFIPEGLLPVTLSLAMAV 309
      + I  +++ +G+I FL  + V+ P+ +FIFA+G++VA +PEGLLP+VTLSLA +
45   Sbjct: 243 LAARIIGALAVAMGVILFLVNLVIYVRLPLETAFIFAIGLMVANVPEGLLPVTLSLAASA 302

Query: 310  QMAKEHALVKKLSSVETLGSVCSDKTGTLTQNMETVNLHWQNGKSYQVTGLGYAPE 369
      ++MA+E+ALVK+LSSVETLG+T++IC+DKTGTLT+ EMTV +W  K +VTG GY PE
Sbjct: 303 RKMARENALVKRLSSVETLGSTTIICTDKTGTLTRGEMTVRKIWIHPHKVIEVTGSGYRPE 362

50   Query: 370 GQILFEGDNICFGNSDRGDLEKLIRFAHLCSNAQVLPNDDRSTYTVLGDPTACLNVL 429
      GQ LF G+ +  + D +L+ L+R A C+++ ++  + ++VLGD TE  L V
Sbjct: 363 GQFLFRGEPV--SHRDMAECLKLMRAATFCNDSALI---HEEGEWSVLGDSTEGALLVAA 417

55   Query: 430 EKSGINIQENRKFPAPRLKELPFDSVRKRMTTIHSLGGDEKDKKISITKGAPKEILDLSY 489
      EK G + +  K PR+ ELPFDS RK MT+IH  G  K+++ KGAPK+I+ LS+
Sbjct: 418 EKLGFDAEAEKAMPRI TELPFDSRRKSMTSIHEKSG----KRVAYVKGAPKKIIGLSER 473

Query: 490  VLSDGKVIPLNKEERNKIQLANDTFAKDGLRVLAVSYCDIEGFSKEQWTQENLEQHMVFI 549
      + DG+V L+ +E+ +I  +D A GLRVLA +Y ++  E  +E+ +V +
60   Sbjct: 474 ISVDGRVRALHADEKERIIGIHDEMASKGLRVLAFAFYRELPE-DLEVRDPGEVERDILV 532

```


-1491-

Query: 550 GLIAMSDDPPREGVREAIKCHAASIRIIMVTGDYGLTALSIAKNIGIIRNDDAKVISGLE 609
 G+ AM DPPREGV+EA++ C A IRIIM+TGDYGLTA +IA+ IGI+ + ++I G E
 Sbjct: 533 GMAAMHDDPPREGVKEAVEHCKTAGIRIIMITGDYGLTAEAIAREIGIVEG-ECRIIKGKE 591

5 Query: 610 LSEMTDSQLKKELSGE--VVFARVAPEQKYRVVTILQEMGEVAVTGDGVNDAPALKKSD 667
 L ++ D++L+ L+ E ++FAR PE K R+ ++L++ E+VA+TGDGVNDAPAL+K+D
 Sbjct: 592 LDKLKDTELRGILARERNLIFARAVPEHKMRIASVLEDSDEIVAMTGDGVNDAPALRKAD 651

10 Query: 668 IGVAMGVTGTDVAKESADMILTDHDFASIVHAVEEGRAVYQNIKKFLTYIFNSNTPEAVP 727
 IGVAMG +GTDVAKE+AD++L DD+FASIV AV EGR VY+NI+KF+TYIF+ T E VP
 Sbjct: 652 IGVAMG-SGTDVAKEAADIVLADDNFASIVTAVREGRTVYENIRKFTTYIFSHETAIVP 710

15 Query: 728 SAFFLFSKGFIPILPLTVMQILAVDLGTMDLPAIGLGVEPPETDVMNRPPRLTDRLLDKG 787
 F + IPLE+T+MQILA+DLGTD LPAL LG PE+DVM PER ++RLL++
 Sbjct: 711 --FIMMVLFSIPLPITIMQILAIIDLGTDTLPALALGRSLPESDVMKLPAPRPPSERLLNRE 768

20 Query: 788 LLIKSFLWYGTIESVLAMGGFFWAHYLRYGNF---TFFVANGIPYREATTMTLGAIIFSQ 844
 ++++ +L+ GTIE+ L M +F Y G + A+ Y ATT+ I+ +Q
 Sbjct: 769 VILRGYLFGTIEAALIMAAYFLVLY--SGGWLPQGEALSASDPLYMRATTVVFAGIVMAQ 826

Query: 845 IGMVMNSRTSYQSIKALSIFGNKLINFGIIMEILAFVLVYVPLFHNLFENTASLGLSHWL 904
 +G +++S+T S + N+ I G++ I L+++Y+P +F TA G+ W
 Sbjct: 827 LGNLLSSQTLRSSALEAGLLRNWILAGMVFAISVMLLVIIYLPPLQPIFGTAPPGILEWF 886

25 Query: 905 YLISCPFIMIGLDEVRLKFSSR 926
 LI I+ DE+RK R
 Sbjct: 887 ILILFTPIVFLTDEMRFIQR 908

There is also homology to SEQ ID 4172.

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1365

A DNA sequence (GBSx1450) was identified in *S. galactiae* <SEQ ID 4173> which encodes the amino acid sequence <SEQ ID 4174>. Analysis of this protein sequence reveals the following:

35 Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.3740(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAB46979 GB:AJ243482 CSRA protein [Enterococcus faecalis]
 Identities = 85/132 (64%), Positives = 105/132 (79%)

Query: 2 KETQEELRQRIGHTAYQVTQNSATEHAFTGKYDDFFEEGIYVDIVSGEVLFSGLDKFQSG 61
 K T+EEL+Q + Y VTQ +ATE F+G+YDDF+++GIYVDIVSGE LFSSLDK+ +G
 Sbjct: 3 KPTEEELKQTLTDLQYAVTQENATERPFSGEYDDFYQDGIYVDIVSGEPLFSGLDKYDAG 62

50 Query: 62 CGWPAFSKPIENRMVTNHQDHSCHGMHRIEVRSRQADSHLGHVFNDDGPVDAGGLRYCINSA 121
 CGWP+F+KPIE R V D SHGMHR+EVRS++ADSHLGHVF DGP+ GGLRYCIN+A
 Sbjct: 63 CGWPSFTKPIEKRGVKEKADFSHGMHREVRSEQADSHLGHVFTDGPLOEGGLRYCINAA 122

55 Query: 122 ALDFIPYDQMAK 133
 AL F+P + K
 Sbjct: 123 ALRFVPVADLEK 134

- 60 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4175> which encodes the amino acid sequence <SEQ ID 4176>. Analysis of this protein sequence reveals the following:

-1492-

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3692(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 109/142 (76%), Positives = 126/142 (87%)

Query: 3 ETQEELRQRIGHTAYQVTQNSATEHAFTGKYDDFFEEGIYVDIVSGEVLFSLLDKFQSGC 62
 ET +EL+QRIG +Y+VTQ++ATE FTG+YD+FFE+GIYVDIVSGEVLFSLLDKF SGC
 Sbjct: 2 ETSDELKQRIGDLSYEVTQHAATESPFTGEYDNFFFEKGIYVDIVSGEVLFSLLDKFN SGC 61

15 Query: 63 GWPAFSKPIENRMVTNHQDHSCHMRHIEVRSRQADSHLGHVFN DGPVDAGGLRYCINSAA 122
 GWPAFSKPIENRMVTNH D S+GM R+EV+SR+A SHLGHVF+DGP +AGGLRYCINSAA
 Sbjct: 62 GWPAFSKPIENRMVTNHDDSSYGMRRVEVKSREAGSHLGHVFS DGPKEAGGLRYCINSAA 121

20 Query: 123 LDFIPYDQMAKRGYGDYLSLFD 144
 L FIPYDQM K GY +L+LFD
 Sbjct: 122 LKFIPYDQMEKEGYAQWLTLFD 143

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 1366

A DNA sequence (GBSx1451) was identified in *S.agalactiae* <SEQ ID 4177> which encodes the amino acid sequence <SEQ ID 4178>. Analysis of this protein sequence reveals the following:

Possible site: 25

30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.1674(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05127 GB:AP001511 unknown [Bacillus halodurans]

40 Identities = 48/152 (31%), Positives = 77/152 (50%), Gaps = 1/152 (0%)

Query: 1 MIRRAKEKDLDPDIAELLKQILMLHHEVRPDIFHTRGSKFSKEQLKEMLIDESKPIFVYES 60
 +IR A +D ++A L Q+ H + R DIF + + + E + V+
 Sbjct: 2 IIREATVQDYEEVARLHTQVHEAHVKERGDIFRSNEPTLNPSFFQAAVQGEKSTVLVFD 61

45 Query: 61 DEGKVVAHLFLQLQEKRDLP- KSFKTLYIDDL CIDEEVRGQQIGQKLMDFARQYAKKHG 119
 + K+ A+ + L + LP + KT+YI DLC+DE RG IG+ ' + + Y K H
 Sbjct: 62 EREKIGAYSVIHLVQTPLEPTMQQRKTVYISDLCVDETRGGGIGRLIFEALISYGAHQ 121

Query: 120 CYNITLNVWNDNQRAVSFYEKLGFKPQQQTQME 151

50 I L+V++ N RA +FY LG + Q+ ME

Sbjct: 122 VDAIELDVYDFNDRAKAFYHSLGMRCQKQTME 153

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 55 vaccines or diagnostics.

-1493-

Example 1367

A DNA sequence (GBSx1452) was identified in *S.agalactiae* <SEQ ID 4179> which encodes the amino acid sequence <SEQ ID 4180>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 52
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3285(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9785> which encodes amino acid sequence <SEQ ID 9786> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:BAB06554 GB:AP001516 unknown conserved protein [Bacillus halodurans]
   Identities = 108/211 (51%), Positives = 149/211 (70%)

   Query: 7  EDVILNATENMVHKLKNDPSGHDWDFHIVRVRNLAVELAHKEGANTFICQMAALLHDIID 66
   E IL + E V +L ++ SGHDW+HI RV +A + +E + F+ Q+AAL HD+ID
20  Sbjct: 3  EQAILQSAEAWVKQLMDEYSGHDWYHIRRVTLMAKAIGEQEKVDFVQVQIAALFHDLDID 62

   Query: 67  DKICQDSKQASYELTQWLYSQDLAIAEVEHILDILENISFKAGTGLTMKTLEGQIVQDAD 126
   DK+ D + A +L W+ + + +++H +DI+ ISFK G G ++ T E +VQDAD
25  Sbjct: 63  DKLVDDPETAKQQLIDWMEAAGVPSQKIDHTMDIINTISFKGGHGQSLATREAMVVQDAD 122

   Query: 127 RLDAMGAIGIARTMAYSGSKGRLIHDPNLKPRENLLEEYRNGQDTAIHFYEKLLKLD 186
   RLDA+GAIGIART AYSG+KG+ I+DP L RE +T+EEYR+G+ TAI HFYEKL KLKD
30  Sbjct: 123 RLDALGAIGIARTFAYSGNKGQPIYDPELPFIRETMTVEEYRHGKSTAINHFYEKLFKLKD 182

   Query: 187 LMNTKQGKMLAQKRHDFLELYLAEFYAEWNG 217
   LMNT+ GK LA++RH F+E ++ F +EWNG
   Sbjct: 183 LMNTETGQLAKERHVFMEQFIERFLSEWNG 213

```

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1368

A DNA sequence (GBSx1453) was identified in *S.agalactiae* <SEQ ID 4181> which encodes the amino acid sequence <SEQ ID 4182>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 21
   >>> May be a lipoprotein

   ----- Final Results -----
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50  !GB:U25448 internalin [Listeria monocytogenes]
   !GB:U25448 internalin [Listeria monocytogenes]
   !GB:U25448 internalin [Listeria monocytogenes]
   !GB:U25448 internalin [Listeria monocytogenes]

   >GP:AAA69530 GB:U25448 internalin [Listeria monocytogenes]
55  Identities = 78/253 (30%), Positives = 132/253 (51%), Gaps = 2/253 (0%)

```

-1494-

Query: 531 LKQLWMTNTGITDYSFLDKMPLLEGLDISQNGIKDLSFLTKYKQLSLIAAANNGITSLKP 590
 L Q+ +N +TD + L + L + ++ N I D++ L L+ + NN IT + P
 Sbjct: 26 LTQINFNNQLTDITPLKDLTKLVDILMNNNQIADITPLANLSNLTGLTLFNNQITDIDP 85

Query: 591 LAELPNLQFLVLSHNNISDLTPLSNLTQELYLHDHNNVKNLSALSGKKDLKVLDSLNNK 650
 L L NL L LS N ISD++ LS LT LQ+L L N V +L L+ L+ LD+S+NK
 Sbjct: 86 LKNLTNLNRLELSSNTISDISALSGLTSLQQLSLG-NQVTDLKPLANLTTLERLDISSNK 144

Query: 651 SADLSTL-KTTSLETLLNETNTSNLSFLKQNPVSNLTINNAKLASLDGIEESDEIVKV 709
 +D+S L K T+LE+L+ S+++ L + L++N +L + + + +
 Sbjct: 145 VSDISVLAKLTNLES LIATNNQISDITPLGILTNLDEL SLNGNQLKDITLASLTNLTDL 204

Query: 710 EAEGNQIKSLVLKKNQGSGLKFLNVTNNQLTSLEGVNNYTSLETLSVSKNKLESLDIKTPN 769
 + NQI +L L L + NQ++++ + T+L L +++N+LE + +
 Sbjct: 205 DLANNQISNLAPLPGLTKLTELKLGANQISNIXPLAGLTALTNLLENQLEDISPISNL 264

Query: 770 KTVTNLDFSHNNV 782
 K +T L NN+
 Sbjct: 265 KNLTYLTYLFNNI 277
 Identities = 91/300 (30%), Positives = 141/300 (46%), Gaps = 42/300 (14%)

Query: 519 INDMTPVLQFKKLKQLWMTNTGITDYSFLDKMPLLEGLDISQNGIKD---LSFLTKYKQL 575
 I D+TP+ L L + N ITD L + L L++S N I D LS LT +QL
 Sbjct: 58 IADITPLANLSNLTGLTLFNNQITDIDPLKNLTNLNRLELSSNTISDISALSGLTSLQQL 117

Query: 576 SLIAAANNGITSLKPLA-----ELPNLQFLVLSHNNISDLTPL 613
 SL N +T LKPLA +L NL+ L+ ++N ISD+TPL
 Sbjct: 118 SL----GNQVTDLKPLANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPL 173

Query: 614 SNLTQELYLHDHNNVKNLSALSGKKDLKVLDSLNNKSADLSTLK-TTSLETLLNETNT 672
 LT L EL L+ N +K++ L+ +L LDL+NN+ ++L+ L T L L L
 Sbjct: 174 GILTNLDEL SLNGNQLKDITLASLTNLTDLDLANNQISNLAPLPGLTKLTELKLGANQI 233

Query: 673 SNLSFLKQNPVSNLTINNAKLASLDGIEESDEIVKVEAGNQIKSLVLKKNQGSGLKFLN 732
 SN+ L ++NL +N +L + I + + N I + + L+ L
 Sbjct: 234 SNIXPLAGLTALTNLLENQLEDISPISNLKNLTLYLTYLFNNISDISPVSSLTKLQRLF 293

Query: 733 VTNNQLTSLEGVNNYTSLETLSVSKNKLESLDIKTPNKTVTNLDHNNVPTSQKLNEK 792
 NN+++ + + N T++ LS N++ L TP +T + +QL LN++
 Sbjct: 294 FYNNKVSVDVSSLANLTNINWLSAGHNQISDL---TPLANLTRI-----TQLGLNDQ 341
 Identities = 73/253 (28%), Positives = 124/253 (48%), Gaps = 4/253 (1%)

Query: 540 GITDYSFLDKMPLLEGLDISQNGIKDLSFLTKYKQLSLIAAANNGITSLKPLAELPNLQF 599
 GI L+ + L ++ S N + D++ L +L I NN I + PLA L NL
 Sbjct: 13 GIKSIDGLEYLNNLTQINFNNQLTDITPLKDLTKLVDILMNNNQIADITPLANLSNLTG 72

Query: 600 LVLSHNNISDLTPLSNLTQELYLHDHNNVKNLSALSGKKDLKVLDSLNNKSADLSTLKT 659
 L L +N I+D+ PL NLT L L L N + ++SALSG L+ L L N +
 Sbjct: 73 LTLFNNQITDIDPLKNLTNLNRLELSSNTISDISALSGLTSLQQLSLGNQVTDLKPLANL 132

Query: 660 TSLETLLNETNTSNLSFLKQNPVSNLTINNAKLASLDGIEESDEIVKVEAGNQIKSL 719
 T+LE L ++ S++S L + + +L N +++ + + + ++ GNQ+K +
 Sbjct: 133 T'LERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPLGILTNLDEL SLNGNQLKDI 192

Query: 720 VLKKNQGSGLKFLNVTNNQLTSLEGVNNYTSLETLSVSKNKLESLDIKTPNKTVTNLDHSH 779
 +L L++ NNQ+++L + T L L + N++ ++ +TNL+ +
 Sbjct: 193 GTLASLTNLTDLDLANNQISNLAPLPGLTKLTELKLGANQISNIXPLAGLTALTNLLENE 252

Query: 780 NNV----PTSQK 788
 N + P S LK
 Sbjct: 253 NQLEDISPISNLK 265
 Identities = 56/209 (26%), Positives = 115/209 (54%), Gaps = 2/209 (0%)

Query: 575 LSLIAAANNGITSLKPLAELPNLQFLVLSHNNISDLTPLSNLTQELYLHDHNNVKNLSA 634
 ++ + A GI S+ L L NL + S+N ++D+TPL +LTKL ++ +++N + +++
 Sbjct: 4 VTTLQADRLGKSIDGLEYLNNLTQINFNNQLTDITPLKDLTKLVDILMNNNQIADITP 63

-1495-

Query: 635 LSGKKDLKVLDSLNNKSADLSTLKT-TSLETLILLNETNTSNLSFLKQNPKVSNTLTINNAK 693
 L+ +L L L NN+ D+ LK T+L L L+ S++S L + L++ N +
 Sbjct: 64 LANLSNLTGLTLFNNQITDIDPLKNLTNLNRLELSSNTISDISALSGLTSLQQLSLGN-Q 122

5 Query: 694 IASLDGIEESDEIVKVEAEGNQIKSLVLKKNQGSGLKFLNVTNNQITSLEGVNNYTSLETL 753
 + L + + +++ N++ + + K +L+ L TNNQ++ + + T+L+ L
 Sbjct: 123 VTDLKLPLANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPLGILTNLDEL 182

10 Query: 754 SVSKNKLESLDIKTPNKTVTNLD FSHNNV 782
 S++ N+L+ + +T+LD ++N +
 Sbjct: 183 SLNGNQLKDIGTLASLTNLTDLDLANNQI 211
 Identities = 61/228 (26%), Positives = 118/228 (51%), Gaps = 3/228 (1%)

15 Query: 483 LATVTKINIGORTNPFQRFGLSLMPNIEVLGIGFTPINDMTPVLQFKKLKQLWMTNTGIT 542
 L ++ ++++G + + L+ + +E L I ++D++ + + L+ L TN I+
 Sbjct: 111 LTSLQQLSLGNQVTDLKP--LANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQIS 168

20 Query: 543 DYSFLDKMPLLEGLDISQNGIKDLSFLT KYKQLSLIAAANNGITSLKPLAELPNLQFLVL 602
 D + L + L+ L ++ N +KD+ L L+ + ANN I++L PL L L L L
 Sbjct: 169 DITPLGILTNLDEL SLNGNQLKDIGTLASLTNLTDLDLANNQISNLAPLPGLTKLTELKL 228

25 Query: 603 SHNNISDLTPLSNLTKLQELYLDHNNVNKLSALSGKKDLKVLDSLNNKSADLSTLKT-TS 661
 N IS++ PL+ LT L L L+ N ++++S +S K+L L L N +D+S + + T
 Sbjct: 229 GANQISNIXPLAGLTALTNLLENQLEDISPISNLKNTLYLTLYFNNISDISPVSSLT K 288

30 Query: 662 LETLLILLNETNTSNLSFLKQNPKVSNTLTINNAKLASLDGIEESDEIVKV 709
 L+ L S++S L ++ L+ + +++ L + I ++
 Sbjct: 289 LQRLFFYNKVS DVSSLANLTNINWLSAGHNQISDLTPLANLTRITQL 336
 Identities = 60/286 (20%), Positives = 129/286 (44%), Gaps = 24/286 (8%)

35 Query: 369 SNKLSDEDQKLIYLAEKLG LNPQNIEVLTS EDGSIIFKYPHDDHSHTIASKDIEIGKPI 428
 +N+++D D K + +L L+ N I +++ G + + + +G +
 Sbjct: 77 NNQITDIDPLKNLTNLNRLELSSNTISDISALSG-----LTSLQQLSLGNQV 123

40 Query: 429 PDGHHDSHSHAKDKVGMATLKGIGFDEIIQDILHADAPTFFPSNETNPEKMRQW--LATV 486
 D K + TL+++ + DI T S ++ L +
 Sbjct: 124 TD-----LKPLANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPLGIL 176

45 Query: 487 TKIN-IGORTNPFQRFGLSLMPNIEVLGIGFTPINDMTPVLQFKKLKQLWMTNTGITDY 544
 T ++ + N + G L+ + N+ L + I+++ P+ KL +L + I++
 Sbjct: 177 TNLDEL SLNGNQLKDIGTLASLTNLTDLDLANNQISNLAPLPGLTKLTELKLGANQISNI 236

50 Query: 545 SFLDKMPLLEGLDISQNGIKDLSFLT KYKQLSLIAAANNGITSLKPLAELPNLQFLVL 604
 L + L L+++N ++D+S ++ K L+ + N I+ + P++ L LQ L +
 Sbjct: 237 XPLAGLTALTNLLENQLEDISPISNLKNTLYLTLYFNNISDISPVSSLT KLQRLFFYN 296

55 Query: 605 NNISDLTPLSNLTKLQELYLDHNNVNKLSALSGKKDLKVLDSLNNK 650
 N +SD++ L+NLT + L HN + +L+ L+ + L L++ +
 Sbjct: 297 NKVSDVSSLANLTNINWLSAGHNQISDLTPLANLTRITQLGLNDQE 342

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4183> which encodes the amino acid sequence <SEQ ID 4184>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAA69530 GB:U25448 internalin [*Listeria monocytogenes*]
 Identities = 88/279 (31%), Positives = 149/279 (52%), Gaps = 2/279 (0%)

-1496-

Query: 419 LPNLETIGIGFTPIKDISPVLQFKKLQQLMTKTGVTDYRFLDNMPQLEGIDISQNNLKD 478
 L + TL IK I + L Q+ + +TD L ++ +L I ++ N + D
 Sbjct: 1 LDXTVTLQADRLGIKSIDGLEYNLNLQINFNNQLTDITPLKDLTKLVDILMNNNQIAD 60

Query: 479 ISFLSKYKNLTLVAAADNGIEDIRPLGQLPNLKFVLSNNKISDLSPLASLHQLQELHID 538
 I+ L+ NLT + +N I DI PL L NL L LS+N ISD+S L+ L LQ+L +
 Sbjct: 61 ITPLANLSNLTGLTLFNNQITDIDPLKNLTNLNRLELSSNTISDISALSGLTSLQQLSL- 119

Query: 539 NNQITDLSFVSHKESLTVVDLSRNADVDLATL-QAPKLETLMVNDTKVSHLDLFLKNNPNL 597
 NQ+TDL P+++ +L +D+S N D++ L + LE+L+ ++S + L NL
 Sbjct: 120 GNQVTDLPLANLTTLERLIDISSNKVSDISVLAKLTNLES LIATNNQISDITPLGILTNL 179

Query: 598 SLSINRAQLQSLEGIEASSVIVRVEAEGNQIKSLVLKDKQGSLLTFLDVTGNQLTSLEGV 657
 LS+N QL+ + + + + ++ NQI +L LT L + NQ++++ +
 Sbjct: 180 DELSLNGNQLKDIGTLASLTNLTDLDLANNQISNLAFLPGLTKLTELKLGANQISNIXPL 239

Query: 658 NNFTALDILSVSKNQLTNVNLSPKNTVTNIDISHNNIS 696
 TAL L +++NQL +++ K +T + + NNIS
 Sbjct: 240 AGLTALTNLELNENQLEDISPISNLKNLTYLTLFNNIS 278

An alignment of the GAS and GBS proteins is shown below.

Identities = 346/753 (45%), Positives = 472/753 (61%), Gaps = 63/753 (8%)

Query: 187 SRLGNQSNSHYRVNSSK-----IAGLHYPTSNGFLFNARG-IGTTPTGILVEHHNH 237
 SR G SN + SK +AG+ +PT +GF+ I T GI+V+H H
 Sbjct: 38 SRGMTSNKIKPIKSKKTKNTHKGVAGVDFPTDDGFILTKDSKILSKTDQGI VVDHGH 97

Query: 238 LHFISFADLRKGGW-----GSIADRYQPQKADSKKQSPSSKKPRTE TLPKDI--KDK 289
 HFI +ADL+ + G+ + ++A S+ S + P DI +D
 Sbjct: 98 SHFIFYADLKGSFPFEYLIPKGASLAKPAVAQRAASQGT SKVADPHHHYEFNPADIVAEDA 157

Query: 290 LAYLARE---LHLDI-----SRIRVLKTLNGEIGFEYPHDDHT 324
 L Y R H + S + T NG G +P D
 Sbjct: 158 LGYTVRHHDDHFIYILKSSLSGQTQAQAKQVATRLPQTSSLVSTATANGIPGLHFPTSDGF 217

Query: 325 HVIMAKDIDLSKPIPNPHHDDH-----HKGHHHD---ESDHKHEEHEHTK 368
 + ++K HD H H +D +++ E H+ +
 Sbjct: 218 QFNQGQIGVGT KD SILVDHGHLPISFADLRQGGWAHVADQYDPAKKAEPETHQTP 277

Query: 369 SNKLSDEDQKLIYLAELGLNPNQIEVLTSDEGSIIFKYPHDDHSHTIASKDIEIGKPI 428
 ++ E Q+KL YLAELG++P+ I+ + ++DG + +YPH DH+H + DIEIGK I
 Sbjct: 278 LSEREKEYQEKLAYLAELGIDPSTIKRVETQDGKLGLEYPHHDAHVMMLSDIEIGKDI 337

Query: 429 PDGH---HDHSHAKDKVGMATLKQIGFDDEIIQDILHA-DAPTFFPSNETNPEKMRQWLA 484
 PD H H K KVG M TL+ +GFD+E+I DI+ DAPTFFPSNE +P M++WLA
 Sbjct: 338 PDPHAIEHARELEKHKVGM DTLRALGFDEEVILDIVRTHDAPTFFPSNEKDPNMMEWLA 397

Query: 485 TVTKINIGQRTNPFQRFGLSLMPNIEVLGIGFTPINDMTPVLQFKKLQQLWMTNTGITDY 544
 TV K+++G R +P QR GLSL+PN+E LGIGFTPI D++PVLQFKKLQQL MT TG+TDY
 Sbjct: 398 TVIKLDLGRKDP LQRKGLSLLPNLETIGIGFTPIKDISPVLQFKKLQQLMTKTGVTDY 457

Query: 545 SFLDKMPLLEGLDISQNGIKDLSFLT KYQLSLIAAANNGITSILKPLAELPNLQFLVLSH 604
 FLD MP LEG+DISQ +KD+SFL+KYK L+L+AAA+NGI ++PL +LPNL+FLVLS+
 Sbjct: 458 RFLDNMPQLEGIDISQNNLKDIFLSKYKNLTLVAAADNGIEDIRPLGQLPNLKFVLSN 517

Query: 605 NNISDLTPLSNLT KLQELYLDHNNVKNLSALSGKKDLKVLDSLNNKSADLSTLKTTSLET 664
 N ISDL+PL++L +LQEL++D+N + +LS +S K+ L V+DLS N DL+TL+ LET
 Sbjct: 518 NKISDLSPLASLHQLQELHIDNNQITDLSFVSHKESLTVVDLSRNADVDLATLQAPKLET 577

Query: 665 LLLNETNTSNLSFLKQNPVSNLTINNKLASLDGIEESDEIVKVEAEGNQIKSLVLKKN 724
 L++N+T S+L FLK NP +S+L+IN A+L SL+GIE S IV+VEAEGNQIKSLVLK+K
 Sbjct: 578 LMVNDTKVSHLDLFLKNNPNLSSLSINRAQLQSLEGIEASSVIVRVEAEGNQIKSLVLKDK 637

Query: 725 QGSLKFLNVTNNQLTSLEGVNNYSLETLSVSKNKLES LDIKTPNKTVTNLDFSHNNVPT 784
 QGSL FL+VT NQLTSLEGVNN+T+L+ LSVSKN+L ++++ PNKTVTN+D SHNN+
 Sbjct: 638 QGSLTFLDVTGNQLTSLEGVNNFTALDILSVSKNQLTNVNLSPKNTVTNIDISHNNISL 697

-1497-

Query: 785 SQLKLNKNIPEAVAKNFPVVEGSMVGNGLAEKAAMASKEDKQVSD-NTNHQKNTKES 843
 + LKLN++IPEA+AKNFPVAV EGSMVGN+ EKAAMA+K + + + +H N +
 Sbjct: 698 ADLKLNEQHIPEATAKNFPVVEGSMVGNGTAEKAAMATKAKESAQEASESHDYNHNHT 757

5 Query: 844 AQANADSKKENPKTHDEHHDHEETDHAHVGH 876
 + E+ D H+HE+ + A +H
 Sbjct: 758 YEDEEGHAHEHRDKDDHDEHEDENEAKDEQNH 790

10 SEQ ID 4182 (GBS84) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 9; MW 97.6kDa).

GBS84-His was purified as shown in Figure 194, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1369

15 A DNA sequence (GBSx1454) was identified in *S.galactiae* <SEQ ID 4185> which encodes the amino acid sequence <SEQ ID 4186>. This protein is predicted to be GTP-binding protein lepa (lepA). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1962(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14493 GB:Z99117 GTP-binding protein [Bacillus subtilis]
 Identities = 464/603 (76%), Positives = 540/603 (88%)

30 Query: 8 KRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQALLDSMDLERERGITIKLNA 67
 +RQ +IRNFSIIAHIDHGKSTLADRILEKT ++ REM+ QLLDSMDLERERGITIKLN+
 Sbjct: 9 ERQSRIRNFSIIAHIDHGKSTLADRILEKTSAITQREMKEQLLDSMDLERERGITIKLNS 68

35 Query: 68 IELNYTAKDGETYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQTLANVYL 127
 ++L Y AKDGE YIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQTLANVYL
 Sbjct: 69 VQLKYKADGEEYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQTLANVYL 128

40 Query: 128 ALDNDLEILPVINKIDLPAADPERVRAEVEDVIGLDASEAVLASAKAGIGIEEILEQIVE 187
 ALDNDLEILPVINKIDLPA+PERVR EVEDVIGLDASEAVLASAKAGIGIEEILEQIVE
 Sbjct: 129 ALDNDLEILPVINKIDLPSAEPERVRQEVEDVIGLDASEAVLASAKAGIGIEEILEQIVE 188

45 Query: 188 KVPAPTGEVDAPLQALIFDSVYDAYRGVILQVRIVNGMVKPGDKIQMMSNGKTFDVTEVG 247
 KVPAPTG+ +APL+ALIFDS+YDAYRGV+ +R+V G VKPG KI+MM+ GK F+VTEVG
 Sbjct: 189 KVPAPTGDPEAPLKALIFDSLYDAYRGVVAYIRVVEGTVPKGQKIKMMATGKEFEVTEVG 248

50 Query: 248 IFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDITITLANNPAIEPLHGYKQMNPMVFAG 307
 +FTPKA + L GDVG++ ASIK V DTRVGDITIT A NPA E L GY+++NPMV+ G
 Sbjct: 249 VFTPKATPTNELTVGDVGFLTASIKNVGDTRVGDITITSAANPAEEALPGYRKLNPVMVYCG 308

55 Query: 368 FNIDLIMTAPSVVYHVNTTDGEMLEVSNPSEFPDPTRVDSIEEPYVKAQIMVPQEFVGVAV 427
 FNIDLITAPSV+Y V TDGE + V NPS PDP +++ +EPPYVKA +MVP ++VGVAV
 Sbjct: 369 FNIDLITAPSVIYDVYMTDGEKVVVDNPSNMPDPQKIERVEEPPYVKAQIMVPNDYVGVAV 428

Query: 428 MELAQKRKRGDFVTMDYIDNVRNVYQIPLAEIVDFDFDKLSSTRGYASFDEISEYRR 487

-1498-

MEL Q KRG+F+ M Y+D NRV++IY +PLAEIV++FFD+LKSST+GYASFDYE+ Y+
 Sbjct: 429 MELCQGRGNFIDMQYLDANRVSIYDMPLAEIVYEFFDQLKSSTKGYASFDYELIGYKP 488

Query: 488 SQLXKMDILLNGDKVDALSFIHVHKEFAYERGKLIVDKLKKIIPRQQFEVPTQAAIGQKIV 547
 S+L KMDI+LNG+K+DALSFIVH++AYERGK+IV+KLK++IPRQQFEVP+QAAIGQKIV

Sbjct: 489 SKLVKMDIMLNGEKIDALSFIHVHRYAYERGKVIVEKLKELIPRQQFEVPTQAAIGQKIV 548

Query: 548 ARSDIKALRKNVLAKCYGGDVSRKRKLEKQKAGKRMKAIGSVEVPQEAFLSVLSMDDD 607
 ARS IKA+RKNVLAKCYGGD+SRKRKLEKQK GK+RMK +GSVEVPQEAFLSVLSMDDD

Sbjct: 549 ARSTIKAMRKNVLAKCYGGDISRKRKLEKQKAGKRMKQVGSVEVPQEAFLSVLSMDDD 608

Query: 608 DKK 610

KK

Sbjct: 609 PKK 611

A related GBS sequence was identified <SEQ ID 10775> which encodes the amino acid sequence <SEQ ID 10776>. A further related GBS nucleic acid sequence <SEQ ID 10955> which encodes amino acid sequence <SEQ ID 10956> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4187> which encodes the amino acid sequence <SEQ ID 4188>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1829(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14493 GB:Z99117 GTP-binding protein [Bacillus subtilis]
 Identities = 463/603 (76%), Positives = 542/603 (89%)

Query: 8 KRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQAQLDSDMLERERGITIKLNA 67
 +RQ +IRNFSIIAHIDHGKSTLADRILEKT ++ REM+ QLLDSDMLERERGITIKLN+

Sbjct: 9 ERQSRIRNFSIIAHIDHGKSTLADRILEKTSAITQREMKEQLDSDMLERERGITIKLNS 68

Query: 68 IELNYTAKDGETYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAQAGIEAQTLANVYL 127
 ++L Y AKDGE YIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAQAGIEAQTLANVYL

Sbjct: 69 VQLKYKAKDGEEYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAQAGIEAQTLANVYL 128

Query: 128 ALDNDLEILPVIKIDLPADPERVRHEVEDVIGLDASEAVLASAKAGIGIEEILEQIVE 187
 ALDNDLEILPVIKIDLP+A+PERVR EVEDVIGLDASEAVLASAKAGIGIEEILEQIVE

Sbjct: 129 ALDNDLEILPVIKIDLPSPERVRQEVEDVIGLDASEAVLASAKAGIGIEEILEQIVE 188

Query: 188 KVPAPTGDVDAPLQALIFDSVYDAYRGVILQVRIVNGIVKPGDKIQMMSGKTFDVTVEVG 247
 KVPAPTGD +APL+ALIFDS+YDAYRGV+ +R+V G VKPG KI+MM+ GK F+VTEVG

Sbjct: 189 KVPAPTGDPEAPLKALIFDSLYDAYRGVVAYIRVVEGTVKPGQKIKMMATGKEFEVTEVG 248

Query: 248 IFTPKAVGRDFLATGDVGVAASIKTVADTRVGDTVTLANNPAKEALHGYKQNMVPMVFAG 307
 +FTPKA +L GDVG++ ASIK V DTRVGDT+T A NPA+EAL GY+++NPMV+ G

Sbjct: 249 VFTPKATPTNELTVGDVGFLTASIKNVGDTRVGDTTITSAANPAEEALPGYRKLNPVMVYCG 308

Query: 308 IYPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFCGFLGLHMDVIQERLERE 367
 +YPI++ KYNDLREALEKL+LND+SLQ+E ETSQALGFGFCGFLG+LHM++IQER+ERE

Sbjct: 309 LYPIDTAKYNDLREALEKLENDSSLYEAETSQALGFGFCGFLGMLHMEIIQERIERE 368

Query: 368 FNIDLIMTAPSVVYHVHTTDEDMEVSNPSEFPDPTRVAFIEEPYVKAQIMVPQEFVGVAV 427
 FNIDLI TAPSV+Y V+ TD + + V NPS PDP ++ +EEPYVKA +MVP ++VGAV

Sbjct: 369 FNIDLITAPSVIYDVYMTDGEKVVVDNPSNMPDPQKIERVEEPYVKAQIMVPQEFVGVAV 428

Query: 428 MELSQKRGDFVTMDYIDNVRNVIIYQIPLAEIVDFDFDKLSSTRGYASFDYDMSEYRR 487

-1499-

MEL Q KRG+F+ M Y+D NRV++IY +PLAEIV++FFD+LKSST+GYASFDY++ Y+
 Sbjct: 429 MELCQKGRGNFIDMQYLDANRVSIIYDMPLEIVYEFFDQLKSSTKGYASFDYELIGYKP 488

Query: 488 SQLVKMDILLNGDKVDALSFIHKEFAYERGKIIVEKLKKIIPRQQFEVPIQAAIGQKIV 547
 S+LVKMDI+LNG+K+DALSFIVH+++AYERGK+IVEKLK++IPRQQFEVP+QAAIGQKIV
 Sbjct: 489 SKLVKMDIMLNGEKIDALSFIHRYDAYERGVIVEKLKELIPRQQFEVPVQAAIGQKIV 548

Query: 548 ARSDIKALRKNVLAKCYGGDVSRKRKLEKQKAGKKRMKAIGSVEVPQEAFLSVLSMDDD 607
 ARS IKA+RKNVLAKCYGGD+SRKRKLEKQK GK+RMK +GSVEVPQEAFF++VL MDD
 Sbjct: 549 ARSTIKAMRKNVLAKCYGGDISRKRKLEKQKEGKRRMKQVGSVEVPQEAFAVLKMDDS 608

Query: 608 TTK 610
 KK
 Sbjct: 609 PKK 611

An alignment of the GAS and GBS proteins is shown below.

Identities = 587/610 (96%), Positives = 601/610 (98%)

Query: 1 MNIEDLKKRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQAQLDSDMLERER 60
 MN +DLKKRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQAQLDSDMLERER
 Sbjct: 1 MNSQLKKRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQAQLDSDMLERER 60

Query: 61 ITIKLNAIELNYTAKDGETYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQ 120
 ITIKLNAIELNYTAKDGETYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQ
 Sbjct: 61 ITIKLNAIELNYTAKDGETYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQ 120

Query: 121 TLANVYLALDNDLEILPVINKIDLPAADPERVRAEVEDVIGLDASEAVLASAKAGIGIEE 180
 TLANVYLALDNDLEILPVINKIDLPAADPERV EVEDVIGLDASEAVLASAKAGIGIEE
 Sbjct: 121 TLANVYLALDNDLEILPVINKIDLPAADPERVRHEVEDVIGLDASEAVLASAKAGIGIEE 180

Query: 181 ILEQIVEKVPAPTGEVDAPLQALIFDSVYDAYRGVILQVRIVNGMVKPGDKIQMMSGKT 240
 ILEQIVEKVPAPTG+VDAPLQALIFDSVYDAYRGVILQVRIVNG+VKPGDKIQMMSGKT
 Sbjct: 181 ILEQIVEKVPAPTGDVDAPLQALIFDSVYDAYRGVILQVRIVNGIVKPGDKIQMMSGKT 240

Query: 241 FDVTEVGIFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDITITLANNPAIEPLHGYKQM 300
 FDVTEVGIFTPKAVGRDFLATGDVG+Y+AAIKTVADTRVGD+TLANNPA E LHGYKQM
 Sbjct: 241 FDVTEVGIFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDITITLANNPAKEALHGYKQM 300

Query: 301 NPMVFAGLYPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFRCGFLGLLHMDVI 360
 NPMVFAG+YPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFRCGFLGLLHMDVI
 Sbjct: 301 NPMVFAGIYPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFRCGFLGLLHMDVI 360

Query: 361 QERLEREFNIDLIMTAPSVVYHVNTDGEMLVSNPSEFPDPTRVDSIEEPYVKAQIMVP 420
 QERLEREFNIDLIMTAPSVVYHV+TTD +M+EVSNPSEFPDPTRV IIEPYVKAQIMVP
 Sbjct: 361 QERLEREFNIDLIMTAPSVVYHVHTDDEDMIEVSNPSEFPDPTRVAFIEEPYVKAQIMVP 420

Query: 421 QEFVGAVMELAQRKRGDFVTMDYIDNVRNVIIYQIPLAEIVDFDFFDKLSSTRGYASFDY 480
 QEFVGAVMEL+QRKRGDFVTMDYIDNVRNVIIYQIPLAEIVDFDFFDKLSSTRGYASFDY
 Sbjct: 421 QEFVGAVMELSQRKRGDFVTMDYIDNVRNVIIYQIPLAEIVDFDFFDKLSSTRGYASFDY 480

Query: 481 EISEYRRSQLKMDILLNGDKVDALSFIHKEFAYERGKLIIVDKLKKIIPRQQFEVPIQA 540
 ++SEYRRSQL KMDILLNGDKVDALSFIHKEFAYERGK+IV+KLKKIIPRQQFEVPIQA
 Sbjct: 481 DMSEYRRSQLVKMDILLNGDKVDALSFIHKEFAYERGKIIVEKLKKIIPRQQFEVPIQA 540

Query: 541 AIGQKIVARSDIKALRKNVLAKCYGGDVSRKRKLEKQKAGKKRMKAIGSVEVPQEAFLS 600
 AIGQKIVARSDIKALRKNVLAKCYGGDVSRKRKLEKQKAGKKRMKAIGSVEVPQEAFLS
 Sbjct: 541 AIGQKIVARSDIKALRKNVLAKCYGGDVSRKRKLEKQKAGKKRMKAIGSVEVPQEAFLS 600

Query: 601 VLSMDDDDKK 610
 VLSMDDD KK
 Sbjct: 601 VLSMDDDTKK 610

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1500-

Example 1370

A DNA sequence (GBSx1455) was identified in *S.agalactiae* <SEQ ID 4189> which encodes the amino acid sequence <SEQ ID 4190>. This protein is predicted to be awd gene product (ndk). Analysis of this protein sequence reveals the following:

```

5   Possible site: 42
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10   bacterial cytoplasm --- Certainty=0.2097(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAF57188 GB:AE003779 awd gene product [Drosophila melanogaster]
    Identities = 73/136 (53%), Positives = 100/136 (72%), Gaps = 5/136 (3%)

    Query: 2   EQTFFMIKPDGVKRGFIGEVISRIERRGFSIDRLEVRYADADILKRHYAELTDRPFFPTL 61
              E+TF M+KPDGV+RG +G++I R E++GF + L+ +A ++L++HYA+L+ RPPFP L
    Sbjct: 25  ERTFIMVKPDGVQRLVGKIIERFEQKGFKLVALKFTWASKELLEKHYADLSARPPFPGL 84

20   Query: 62  VDYMTSGPVIIGVISGEEVISTWRTMMGSTNPKDALPGTIRGDFQAQAPSPNQATCNIVHG 121
              V+YM SGPV+ V G V+ T R M+G+TNP D+LPGTIRGDF Q NI+HG
    Sbjct: 85  VNYMNSGPVVPVMVWGLNVVKTGRQMLGATNPADSLPGTIRGDFC-----IQVGRNIHG 139

25   Query: 122 SDSPESATREIAIWFN 137
              SD+ ESA +EIA+WFN
    Sbjct: 140 SDAVESAEKEIALWFN 155

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4191> which encodes the amino acid sequence <SEQ ID 4192>. Analysis of this protein sequence reveals the following:

```

30   Possible site: 22
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
35   bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40   Identities = 30/48 (62%), Positives = 35/48 (72%)

    Query: 87  MMGSTNPKDALPGTIRGDFQAQAPSPNQATCNIVHGS DSPESATREIAI 134
              MM TNPKDAL GTIR +FAQAP + N+VHGS S +SA REIA+
    Sbjct: 1   MMRVTNPKDALCGTIRENFAQAPGDDGGIFNMVHGSHSRDSARREIAL 48

45

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1371

A DNA sequence (GBSx1456) was identified in *S.agalactiae* <SEQ ID 4193> which encodes the amino acid sequence <SEQ ID 4194>. Analysis of this protein sequence reveals the following:

```

50   Possible site: 15
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
55   bacterial cytoplasm --- Certainty=0.2734(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```